

Data-Driven Time-Frequency Analysis

Thomas Y. Hou*

Zuoqiang Shi†

February 28, 2012

Abstract

In this paper, we introduce a new adaptive data analysis method to study trend and instantaneous frequency of nonlinear and non-stationary data. This method is inspired by the Empirical Mode Decomposition method (EMD) and the recently developed compressed (compressive) sensing theory. The main idea is to look for the sparsest representation of multiscale data within the largest possible dictionary consisting of intrinsic mode functions of the form $\{a(t) \cos(\theta(t))\}$, where $a \in V(\theta)$, $V(\theta)$ consists of the functions smoother than $\cos(\theta(t))$ and $\theta' \geq 0$. This problem can be formulated as a nonlinear L^0 optimization problem. In order to solve this optimization problem, we propose a nonlinear matching pursuit method by generalizing the classical matching pursuit for the L^0 optimization problem. One important advantage of this nonlinear matching pursuit method is it can be implemented very efficiently and is very stable to noise. Further, we provide a convergence analysis of our nonlinear matching pursuit method under certain scale separation assumptions. Extensive numerical examples will be given to demonstrate the robustness of our method and comparison will be made with the EMD/EEMD method. We also apply our method to study data without scale separation, data with intra-wave frequency modulation, and data with incomplete or under-sampled data.

1 Introduction

Developing a truly adaptive data analysis method is important for our understanding of many natural phenomena. Traditional data analysis methods, such as the Fourier transform, use pre-determined basis. They provide an effective tool to process linear and stationary data. However, there are still some limitations in applying these methods to analyze nonlinear and nonstationary data. Time-frequency analysis has been developed to overcome the limitations of the traditional techniques by representing a signal with a joint function of both time and frequency. The recent advances of wavelet analysis have opened a new path for time-frequency analysis. A significant breakthrough of wavelet analysis is the use

*Applied and Comput. Math, Caltech, Pasadena, CA 91125. *Email: hou@cms.caltech.edu.*

†Mathematical Sciences Center, Tsinghua University, Beijing, China, 100084. *Email: zqshi@math.tsinghua.edu.cn.*

of multi-scales to characterize signals. This technique has led to the development of several wavelet-based time-frequency analysis techniques [15, 6, 19].

Another important approach in the time-frequency analysis is to study instantaneous frequency of a signal. Some of the pioneering work in this area was due to Van der Pol [28] and Gabor [11], who introduced the so-called Analytic Signal (AS) method that uses the Hilbert transform to determine instantaneous frequency of a signal. This Analytic Signal method is one of the most popular ways to define instantaneous frequency. Until very recently, this method works mostly for monocomponent signals in which the number of zero-crossings is equal to the number of local extreme [1]. There were other attempts to define instantaneous frequency such as the zero-crossing method [25, 26, 20] and the Wigner-Ville distribution method [1, 17, 23, 9, 16, 22]. However most of these methods are rather restrictive. More substantial progress has been made only recently with the introduction of the EMD method [14]. The EMD method provides an effective tool to decompose a signal into a collection of intrinsic mode functions (IMF) that allow well-behaved Hilbert transforms for computation of physically meaningful time-frequency representation. We remark that the Hilbert spectral representation based on the wavelet projection has also been carried in [21].

Inspired by the EMD method and the recently developed compressed (compressive) sensing theory, we propose a data-driven time-frequency analysis method. There are two important ingredients of this method. The first one is that the basis that is used to decompose the data is derived from the data rather than determined *a priori*. This explains the name “data-driven” in our method. The second ingredient is to look for the sparsest decomposition of the signal among the largest possible dictionary consisting of intrinsic mode functions. The adoption of this data-driven basis and the search for the sparsest decomposition over this highly redundant basis make our time-frequency analysis method fully adaptive to the signal. As we are going to demonstrate later, our method can reveal some hidden physical information of the signal, such as trend and instantaneous frequency.

Our data-driven time-frequency analysis method is motivated by the observation that many multi-scale data often have an intrinsic sparse structure in the time-frequency plane, although its representation in the physical domain could be quite complicated. The challenge is that such sparse representation is valid only for certain multiscale basis that is adapted to the data and is unknown *a priori*. Finding such nonlinear multiscale basis is an essential ingredient of our method. In this sense, our problem is more difficult than the compressed (compressive) sensing problem in which the basis is assumed to be known *a priori*. One way to find the adaptive basis is to learn from the data if we have a large number of data samples that share the similar physical property. This does not apply to our problem since we deal with only a single signal. We overcome this difficulty by reformulating the problem as a nonlinear optimization among the largest possible dictionary. The trade-off is that such decomposition is not unique. We need to exploit the intrinsic sparse structure of the data to select the sparsest one among all the possible decompositions.

In our method, the dictionary is given as following:

$$\mathcal{D} = \{a(t) \cos \theta(t) : \theta'(t) \geq 0, a(t), \theta'(t) \in V(\theta)\}, \quad (1)$$

where $V(\theta)$ is a linear space consisting of functions smoother than $\cos \theta(t)$. The construction of $V(\theta)$ with given $\theta(t)$ is in general an overcomplete Fourier basis given below:

$$V(\theta) = \text{span} \left\{ 1, \cos \left(\frac{k\theta}{2L_\theta} \right), \sin \left(\frac{k\theta}{2L_\theta} \right), k = 1, \dots, 2\lambda L_\theta \right\}, \quad (2)$$

where $L_\theta = \lfloor \frac{\theta(1)-\theta(0)}{2\pi} \rfloor$, $\lfloor \mu \rfloor$ is the largest integer less than μ , and $\lambda \leq 1/2$ is a parameter to control the smoothness of $V(\theta)$. We then decompose the signal over this dictionary by looking for the sparsest decomposition. The sparsest decomposition can be obtained by solving a nonlinear optimization problem:

$$\begin{aligned} P : \quad & \text{Minimize} \quad M \\ & \text{Subject to: } f(t) = \sum_{k=1}^M a_k(t) \cos \theta_k(t), \quad a_k(t) \cos \theta_k(t) \in \mathcal{D}, \quad k = 1, \dots, M. \end{aligned} \quad (3)$$

When the signal is polluted by noise, the equality in the above constraint is relaxed to be an inequality depending on the noise level. This optimization problem can be viewed as a nonlinear version of the L^0 minimization problem and is known to be very challenging to solve. Inspired by the compressed (compressive) sensing theory [2], we propose a l^1 -regularized nonlinear matching pursuit method to solve this nonlinear optimization problem.

Our nonlinear matching pursuit is inspired by the linear matching pursuit method [18, 27]. We first extract an intrinsic mode function $a(t) \cos \theta(t) \in \mathcal{D}$ from the signal $f(t)$ by looking for the one which matches the signal $f(t)$ best among all the elements in \mathcal{D} . This would imply the following nonlinear optimization problem:

$$\text{Minimize } \gamma \|\hat{a}(t)\|_{l^1} + \|f(t) - a(t) \cos \theta(t)\|_{l^2}^2, \quad \text{Subject to } a(t) \cos \theta(t) \in \mathcal{D},$$

where $\gamma > 0$ is a regularization parameter and $\hat{a}(t)$ is the representation of $a(t)$ in $V(\theta)$. In some cases when the signals are periodic, we can choose $\gamma = 0$. Denote by $r(t)$ the residual after subtracting $a(t) \cos \theta(t)$ from $f(t)$, i.e. $r(t) = f(t) - a(t) \cos \theta(t)$. We can then treat $r(t)$ as a new signal to extract the remaining IMFs. There are two important advantages of this nonlinear matching pursuit approach. The first one is that this method is very stable to noise perturbation. The second one is that it can be implemented very efficiently. In the case of $\gamma = 0$, the resulting method can be solved approximately by Fast Fourier Transform, and the complexity of our algorithm is of order $O(N \log N)$ where N is the number of data sample points that we use to represent the signal. The low computational cost and the robustness to noise perturbation make this method very effective in many applications. Moreover, for data that satisfy certain scale separation conditions, we prove that our method recovers the IMFs and their instantaneous frequencies accurately.

We perform extensive numerical experiments to test the robustness and the accuracy of our data-driven time-frequency analysis method for both synthetic data and some real data. Our results show that the nonlinear matching pursuit can indeed decompose a multiscale signal into a sparse collection of intrinsic mode functions. We also compare our method with the original EMD method. For the data without noise, we find that our method gives results comparable to those obtained by the EMD method. Moreover, for noisy data, our method seems to provide better estimation of the instantaneous frequency and IMFs than EMD and recently developed EEMD method [30].

A common difficulty in many data analysis methods is the relatively large error produced near the boundary of the data set. For the EMD method, this source of error is referred to as the “end effect”, which is primarily caused by the use of cubic spline interpolation in constructing the envelope and the median of the signal. Our data-driven time-frequency analysis method seems to be less sensitive to this end effect, especially when the data satisfy certain scale separation property.

We have also extended our data-driven time-frequency analysis method to decompose data that do not have a good scale separation property. By incorporating a shape function into our dictionary that is adapted to the signal, we can also extend our method to decompose data with strong intra-wave modulation. Finally, we demonstrate that our data-driven time-frequency analysis method can be applied to recover the original signal with missing data in certain interval. The recovered signal as well as their instantaneous frequency seems to have reasonably good accuracy. We also apply our method to decompose under-sampled data. The result is quite encouraging even if the under-sampled data are polluted by noise.

We remark that there has been some recent progress in developing a mathematical framework for an EMD like method using synchrosqueezed wavelet transforms by Daubechies, Lu and Wu [7]. This seems to be a very promising approach. We have performed some preliminary numerical experiments to compare the performance of our method with the synchrosqueezed wavelet approach. In many cases, we find that the two methods give comparable and complementary results. We are currently exploring a hybrid approach that combines the advantages of our method with those of the synchrosqueezed wavelet approach. Our preliminary results seem quite encouraging. We will report this in a forthcoming paper.

The remaining of the paper is organized as follows. In Section 2, we give a brief review of some existing data analysis methods such as the matching pursuit, the basis pursuit and the EMD method. We introduce our adaptive data analysis method. In Section 3, a simplified version of our data-driven time-frequency data analysis method is introduced for periodic data which can be implemented efficiently by using the Fast Fourier Transform. In Section 4, we present some numerical experiments, include incomplete or under-sampled data, to demonstrate the performance of our method. In Section 5, we generalize our data-driven time-frequency data analysis method to analyze data with poor scale separation property and data with strong intra-wave modulation. We present some preliminary error analysis of our data-driven time-frequency data analysis method in Section 6. The technical proof of the main result is deferred to the Appendix. Some conclusions are made in Section 7.

2 Brief review of the existing sparse decomposition methods

A considerable focus of activities in the recent signal processing literature has been the development of the sparse signal representations over a redundant dictionary. Among these methods, the matching pursuit [18] and the basis pursuit [5] have attracted a lot of attention in recent years due to the development of the compressed (compressive) sensing. All these methods consist of two parts: a dictionary to decompose the signal and a decomposition method to select the sparsest decomposition.

2.1 Dictionaries

A dictionary is a collection of parameterized waveforms $\mathcal{D} = \{\phi_\gamma\}_{\gamma \in \Gamma}$. Many dictionaries have been proposed in the literature. Here we review a few of them that have been used widely.

A Fourier Dictionary. A Fourier dictionary is a collection of sinusoidal waveforms. More specifically, the waveforms consist of the following two families,

$$\phi_{\omega,0} = \cos(\omega t), \quad \phi_{\omega,1} = \sin(\omega t). \quad (4)$$

For the standard Fourier dictionary, ω runs through the set of all cosines with Fourier frequencies $\omega_k = 2k\pi/n$, $k = 0, 1, \dots, n/2$, and all sines with Fourier frequencies $\omega_k = 2k\pi/n$, $k = 1, \dots, n/2 - 1$, where n is the number of sample points. We can also obtain an overcomplete Fourier dictionary by sampling the frequencies more finely. Let $l > 1$. We may choose $\omega_k = 2k\pi/(ln)$, $k = 0, 1, \dots, ln/2$ for cosines and $\omega_k = 2k\pi/(ln)$, $k = 1, \dots, ln/2 - 1$ for sines. This is an l -fold overcomplete system. In the algorithm for non-periodic data, we will use this kind of overcomplete Fourier dictionary.

A Wavelet Dictionary. A wavelet dictionary is a collection of translations and dilations of the basic mother wavelet ψ , together with translations of the scaling function φ defined below:

$$\phi_{a,b,0} = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \quad \phi_{a,b,1} = \frac{1}{\sqrt{a}}\varphi\left(\frac{t-b}{a}\right). \quad (5)$$

For the standard wavelet dictionary, we let a, b run through the discrete collection of mother wavelets with dyadic scales $a_j = 2^j/n$, $j = j_0, \dots, \log_2(n) - 1$, and locations that are integer multiples of the scale $b_{j,k} = ka_j$, $k = 0, \dots, 2^j - 1$, and the collection of scaling functions at the coarse scale j_0 . This dictionary consists of n waveforms, which form an orthonormal basis. As in the Fourier dictionary, an overcomplete wavelet dictionary can be obtained by sampling the locations more finely.

A Time-Frequency Dictionary. A typical time-frequency dictionary is the Gabor dictionary due to Gabor (1946). In this dictionary, we take $\gamma = (\omega, \tau, \theta, \delta)$, where $\omega \in [0, \pi)$ is

frequency, τ is a location, θ is a phase, and δ is the duration. We define the waveform as follows:

$$\phi_\gamma(t) = \exp\left(-\frac{(t-\tau)^2}{\delta^2}\right) \cos(\omega(t-\tau) + \theta). \quad (6)$$

Such waveforms consist of frequencies near ω and essentially vanish far away from τ .

An EMD Dictionary. We can also define a dictionary via the EMD method. In the EMD method [14], the dictionary is the collection of all Intrinsic Mode Functions (IMF), which are functions defined descriptively by enforcing the following two conditions:

1. The number of the extreme and the number of the zero crossings of the function must be equal or differ at most by one;
2. At any point of the function, the average of the upper envelope and the lower envelope defined by the local extreme should be zero (symmetric with respect to zero).

Inspired by the EMD method, we will use a variant of the EMD dictionary to construct a sparse decomposition of a signal via nonlinear optimization.

2.2 Decomposition Methods

In this subsection, we review a few decomposition methods that can be used to give a sparse decomposition of a signal by exploiting the intrinsic sparsity structure of the signal. In recent years, there have been a lot of research activities in looking for the sparsest representation of a signal over a redundant dictionary [18, 5, 8, 3, 4], i.e. look for a decomposition of a signal f over a given dictionary $\mathcal{D} = \{\phi_\gamma\}_{\gamma \in \Gamma}$ as

$$f = \sum_{k=1}^m \alpha_{\gamma_k} \phi_{\gamma_k} + R^m, \quad (7)$$

with the smallest m , where R^m is the residual. Whether or not a signal can be decomposed into a sparse decomposition depends on the choice of the dictionary that we use to decompose the signal. In general, a more redundant dictionary tends to give better adaptivity, which implies better sparsity of the decomposition. However, when the dictionary is highly redundant, the decompositions are not unique. We need to give a criterion to pick up the “best” decomposition among all the possible choices.

Matching Pursuit. In [18], Mallat and Zhang introduced a general decomposition method called the matching pursuit that exploits the sparsity of a signal. Starting from an initial approximation $\mathbf{s}^0 = 0$ and a residual $\mathbf{r}^0 = \mathbf{s}$, the matching pursuit builds up a sequence of sparse approximations step by step. At stage k , the method identifies an atom that best matches the residual and then adds it to the current approximation, so that $\mathbf{s}^k = \mathbf{s}^{k-1} + \alpha_k \phi_{\gamma_k}$, where $\alpha_k = \langle \mathbf{r}^{k-1}, \phi_{\gamma_k} \rangle$ and $\mathbf{r}^k = \mathbf{s} - \mathbf{s}^k$. After m steps, one has a

representation of the form (7), with residual $R^m = r^m$. A similar algorithm was proposed for Gabor dictionaries by S. Qian and D. Chen [24].

An intrinsic feature of this algorithm is that when stopped after a few steps, it yields an approximate sparse representation using only a few atoms. When the dictionary is orthogonal, the method works perfectly. If the dictionary is not orthogonal, the situation is less clear. Recently, J. Tropp and A. Gilbert proved that under some assumptions on the basis, the orthogonal matching pursuit can solve the original l_0 minimization problem [27].

Basis Pursuit. Another important class of decomposition methods is the basis pursuit, which was introduced by S. Chen, D. Donoho and M. Saunders [5]. First, we reformulate the decomposition problem in the following way. Suppose we have a discrete dictionary of p waveforms and we collect all these waveforms as columns of an n by p matrix Φ . The decomposition problem (7) can be reformulated as:

$$\mathbf{s} = \Phi\alpha, \quad (8)$$

where $\alpha = (\alpha_\gamma)$ is the vector of coefficients in (7).

The basic idea of the basis pursuit is to find a sparse representation of the signal whose coefficients have a minimal l_1 norm, i.e. the decomposition is obtained by solving the problem

$$\min \|\alpha\|_{l_1}, \quad \text{subject to } \Phi\alpha = \mathbf{s}. \quad (9)$$

Recently, the basis pursuit has received a lot of attention, since it is found that under some conditions the basis pursuit can recover the exact solution of the original l_0 minimization problem [3, 8]. There has been extensive research to obtain a sparse representation by the basis pursuit in a variety of applications. An essential component of the basis pursuit is to solve the l^1 minimization problem. The computational cost of solving this l^1 minimization is more expensive than the least-square problem in the matching pursuit, although a powerful Split Bregman method has been introduced by Goldstein and Osher to speed up the l^1 minimization problem considerably [12].

The EMD decomposition via a sifting process. The EMD method decomposes a signal to its IMFs sequentially. The basic idea behind this approach is the removal of the local median from a signal by using a sifting process. Specifically, for a given signal, $f(t)$, one tries to decompose it as a sum of the local median $m(t)$, and an IMF. A cubic spline polynomial is used to interpolate all the local maxima to obtain an upper envelope, and to interpolate all the local minima to obtain a lower envelope. By averaging the upper and lower envelopes, one obtains an approximate median for $m(t)$. One then decides whether or not to accept the obtained $m(t)$ as our local median depending on whether $f(t) - m(t)$ gives an acceptable IMF that satisfies the two conditions that are specified in the definition of an EMD dictionary. If $f(t) - m(t)$ does not satisfy these conditions, one can treat $f(t) - m(t)$ as a new signal and construct a new candidate for the IMF by using the same procedure

described above. This sifting process continues until we obtain a satisfactory IMF, which we denote as $f_n(t)$. Now we can treat $f(t) - f_n(t)$ as a new signal, and apply the same procedure to generate the second IMF, $f_{n-1}(t)$. This procedure continues until $f_0(t)$ is either monotone or contains at most one extremum. For more details of the sifting process, we refer to [14].

Decomposition based on a nonlinear TV^3 minimization. Inspired by the EMD method, we proposed a decomposition method based on a nonlinear TV^3 minimization in our previous paper [13]. Here TV^3 is the total variation of the third order derivative of a function, defined as $TV^3(f) = \int_a^b |f^{(4)}(t)|dt$. We use a TV^3 norm because the L^1 norm or the total variation norm is not strong enough to enforce the regularity of the median or the envelope of our decomposition. For example, the use of the total variation norm, which is very popular in imaging processing community, tends to give a decomposition whose median or envelope is piecewise constant, which is referred to as the stair-case effect. The TV^3 norm, on the other hand, gives a much smoother decomposition for both the median and the envelope. Incidentally, the minimization using the TV^3 norm tends to favor piecewise cubic polynomials such as cubic splines. Thus, our method gives results that are qualitatively similar to those obtained by the EMD method which uses cubic splines to construct its median and envelope from the local extrema of the signal.

We now give a brief review of our TV^3 decomposition method. In our approach, every element in our dictionary automatically satisfies the conditions of IMF. There is no need to do any sifting or use the Hilbert transform in our method. First, we decompose a signal $f(t)$ into its local median a_0 and an IMF $a_1 \cos \theta(t)$ by solving the following nonlinear optimization problem:

$$\begin{aligned} (P) \quad & \text{Minimize} \quad TV^3(a_0) + TV^3(a_1), \\ & \text{Subject to:} \quad a_0(t) + a_1(t) \cos \theta(t) = f(t), \quad \theta'(t) \geq 0. \end{aligned} \quad (10)$$

To solve this nonlinear optimization problem, we proposed the following Newton type of iterative method:

Initialization: $\theta^0 = \theta_0$.

Main Iteration:

Step 1: Update a_0^n , a_1^n , b_1^n by solving the following linear optimization problem:

$$\text{Minimize} \quad TV^3(a_0^n) + TV^3(a_1^n) + TV^3(b_1^n), \quad (11)$$

$$\text{Subject to :} \quad a_0^n + a_1^n \cos \theta^{n-1}(t) + b_1^n \sin \theta^{n-1}(t) = f(t). \quad (12)$$

Step 2: Update the phase function θ :

$$\theta^n = \theta^{n-1} - \mu \arctan \left(\frac{b_1^n}{a_1^n} \right), \quad (13)$$

where $\mu \in [0, 1]$ is chosen to enforce that θ^n is an increasing function:

$$\mu = \max \left\{ \alpha \in [0, 1] : \frac{d}{dt} \left(\theta_k^{n-1} - \alpha \arctan \left(\frac{b_1^n}{a_1^n} \right) \right) \geq 0 \right\}. \quad (14)$$

Step 3: If $\|\theta^n - \theta^{n-1}\|_2 \leq \epsilon_0$, stop. Otherwise, go to Step 1.

In [13], we performed a number of numerical experiments and compared the results with those obtained by the EMD (or EEMD) method. Our results show that this method shares many important properties with the original EMD method. Moreover, its performance does not depend on numerical parameters such as the number of sifting or the stop criterion, which seem to have a major effect on the original EMD method.

There are two limitations of this approach. The first one is that the computational cost to solve the TV^3 minimization problem is relatively high, even if we use the Split Bregman method of Goldstein and Osher [12]. The second one is that this method is more sensitive to noise perturbation, although a nonlinear filter was introduced to alleviate this difficulty. In comparison, the nonlinear matching pursuit method we introduce in this paper is very stable to noise perturbation and has a relatively low computational cost.

3 Sparse time-frequency decomposition method based on nonlinear matching pursuit

Our adaptive data analysis method is based on finding the sparsest decomposition of a signal by solving a nonlinear optimization problem. First, we need to construct a large dictionary that can be used to obtain a sparse decomposition of the signal. In principle, the larger the dictionary is, the more adaptive (or sparser) the decomposition is.

3.1 Dictionary

In our method, the dictionary is chosen to be:

$$\mathcal{D} = \{a(t) \cos \theta(t) : \theta'(t) \geq 0, a(t), \theta'(t) \text{ is smoother than } \cos \theta(t)\}. \quad (15)$$

Let $V(\theta)$ be the collection of all the functions that are smoother than $\cos \theta(t)$. In general, it is most effective to construct $V(\theta)$ as an overcomplete Fourier basis given below:

$$V(\theta) = \text{span} \left\{ 1, \cos \left(\frac{k\theta}{2L_\theta} \right), \sin \left(\frac{k\theta}{2L_\theta} \right), k = 1, \dots, 2\lambda L_\theta \right\}, \quad (16)$$

where $L_\theta = \lfloor \frac{\theta(1)-\theta(0)}{2\pi} \rfloor$, $\lfloor \mu \rfloor$ is the largest integer less than μ , and $\lambda \leq 1/2$ is a parameter to control the smoothness of $V(\theta)$. Then \mathcal{D} can be written as

$$\mathcal{D} = \{a(t) \cos \theta(t) : \theta'(t) \geq 0, a(t), \theta'(t) \in V(\theta)\}. \quad (17)$$

In some sense, the dictionary \mathcal{D} defined above can be considered as a collection of IMFs. This property makes our method as adaptive as the EMD method. We also call an element of \mathcal{D} as an IMF. Since the dictionary \mathcal{D} is highly redundant, the decomposition over this dictionary is not unique. We need a criterion to select the “best” one among all possible decompositions. We assume that the data we consider have an intrinsic sparse structure in the time-frequency plane in some nonlinear and nonstationary basis. However, we do not know this basis *a priori* and we need to derive (or learn) this basis from the data. Based on this consideration, we adopt sparsity as our criterion to choose the best decomposition. This criterion yields the following nonlinear optimization problem:

$$\begin{aligned} P : \quad & \text{Minimize} \quad M \\ & \text{Subject to: } f(t) = \sum_{k=1}^M a_k(t) \cos \theta_k(t), \quad a_k(t) \cos \theta_k(t) \in \mathcal{D}, \quad k = 1, \dots, M, \end{aligned} \quad (18)$$

or

$$\begin{aligned} P_\delta : \quad & \text{Minimize} \quad M \\ & \text{Subject to: } \|f(t) - \sum_{k=1}^M a_k(t) \cos \theta_k(t)\|_{l^2} \leq \delta, \quad a_k(t) \cos \theta_k(t) \in \mathcal{D}, \quad k = 1, \dots, M, \end{aligned} \quad (19)$$

if the signal has noise with noise level δ .

After this optimization problem is solved, we get a very clear time-frequency representation:

$$\text{Frequency: } \omega_k(t) = \theta'_k(t), \quad \text{Amplitude: } a_k(t). \quad (20)$$

3.2 Nonlinear matching pursuit

The above optimization problem can be seen as a nonlinear L^0 minimization problem. Thanks to the recent developments of the compressed (compressive) sensing, it is now well known that there are two kinds of methods to solve a L^0 minimization problem: the basis pursuit and the matching pursuit. Since the dictionary we adopt here has infinitely many elements, the generalization of the basis pursuit is not so straightforward. But the idea of the matching pursuit can be generalized. Applying the idea of the matching pursuit to our problem, we obtain the following algorithm:

- $r^0 = f(t)$.

Step 1: Solve the following l^1 -regularized nonlinear least-square problem (P_2):

$$\begin{aligned} P_2 : \quad & \text{Minimize} \quad \gamma \|\widehat{a}_k\|_{l^1} + \|r^{k-1} - a_k \cos \theta_k\|_{l^2}^2 \\ & \text{Subject to: } \theta'_k \geq 0, \quad a_k(t) \in V(\theta_k), \end{aligned} \quad (21)$$

where $\gamma > 0$ is a regularizing parameter and \widehat{a}_k is the representation of a_k in the $V(\theta_k)$ space.

Step 2: Update the residual

$$r^k = f(t) - \sum_{j=1}^k a_j(t) \cos \theta_j(t). \quad (22)$$

Step 3: If $\|r^k\|_{l^2} < \epsilon_0$, stop. Otherwise, go to Step 1.

In the first step of the above iterative algorithm, unlike the standard matching pursuit, we solve a l^1 regularized nonlinear least-square problem P_2 . To solve this problem, we use the following Gauss-Newton type method:

- $\theta_k^0 = \theta_0$.

Step 1: Solve the following l_1 regularized least-square problem:

$$\begin{aligned} P_{2,l_2} \quad & \text{Minimize} \quad \gamma(\|\widehat{a}_k^{n+1}\|_{l^1} + \|\widehat{b}_k^{n+1}\|_{l^1}) + \|r^{k-1} - a_k^{n+1}(t) \cos \theta_k^n(t) - b_k^{n+1}(t) \sin \theta_k^n(t)\|_{l^2}^2 \\ & \text{Subject to} \quad a_k^{n+1}(t), b_k^{n+1}(t) \in V(\theta_k^n). \end{aligned}$$

where $\widehat{a}_k^{n+1}, \widehat{b}_k^{n+1}$ are the representations of a_k^{n+1}, b_k^{n+1} in the $V(\theta_k^n)$ space.

Step 2: Update θ_k^n ,

$$\theta_k^{n+1} = \theta_k^n - \lambda \arctan \left(\frac{b_k^{n+1}}{a_k^{n+1}} \right), \quad (23)$$

where $\lambda \in [0, 1]$ is chosen to make sure that θ_k^{n+1} is a monotonely increasing function.

$$\lambda = \max \left\{ \alpha \in [0, 1] : \frac{d}{dt} \left(\theta_k^n - \alpha \arctan \left(\frac{b_k^{n+1}}{a_k^{n+1}} \right) \right) \geq 0 \right\}. \quad (24)$$

Step 3: If $\|\theta_k^{n+1} - \theta_k^n\|_{l^2} < \epsilon_0$, stop. Otherwise, go to Step 1.

In the first step of above algorithm, we solve a l_1 regularized least square problem. The l^1 regularization tends to stabilize the least square problem using an overcomplete Fourier basis. It also favors a sparse decomposition of the data.

3.3 A fast algorithm for periodic data

In the iterative algorithm given in previous section, we need to solve a l_1 regularized least square problem in each step. This is the most expensive part of the algorithm especially when the number of the data points is large. In this subsection, we consider the special case when the data are periodic. In this case, we can introduce an method based on Fast Fourier Transform(FFT) instead of solving the l_1 regularized least square problem.

One big advantage for periodic data is that we can use the standard Fourier basis to construct the $V(\theta)$ space in the following way instead of the overcomplete Fourier basis given in (2).

$$V(\theta) = \text{span} \left\{ \cos \left(\frac{k\theta}{L_\theta} \right), \sin \left(\frac{l\theta}{L_\theta} \right) : k = 0, \dots, \lambda L_\theta, l = 1, \dots, \lambda L_\theta \right\}, \quad (25)$$

where $\lambda \leq 1/2$ is a parameter to control the smoothness of functions in $V(\theta)$ and $L_\theta = (\theta(1) - \theta(0))/2\pi$ is a positive integer.

The standard Fourier basis is orthogonal to each other, so the l_1 regularized term is not necessary in this case. Then, the least-square problem that we need to solve in the iterative algorithm is described below (recall that we set $\gamma = 0$):

$$\begin{aligned} & \text{Minimize} \quad \|r(t) - a(t) \cos \theta(t) - b(t) \sin \theta(t)\|_{l_2}^2 \\ & \text{Subject to} \quad a(t), b(t) \in V(\theta). \end{aligned} \quad (26)$$

In order to simplify the notations, we drop the subscript and superscript here.

Notice that in the iterative process, the derivative of the phase function $\theta(t)$ is always monotonically increasing. Thus, we can use $\theta(t)$ as a new coordinate. In this new coordinate, $\cos \theta, \sin \theta$ and the basis of $V(\theta)$ are simple Fourier modes, then the least-square problem can be solved by using the Fourier Transform. More specifically, for a given increasing phase function $\theta(t)$, we have the following algorithm to solve the least-square problem (26) approximately:

Step 1: Apply an interpolation method to obtain the value of $r(t)$ at a uniform mesh in the θ -coordinate, $r_\theta(\theta_j)$:

$$r_\theta(\theta_j) = \text{Interpolate}(\theta(t_i), r, \theta_j), \quad (27)$$

where $\theta_j, j = 1, \dots, N$ are uniformly distributed in the θ -coordinate.

Step 2: Apply the low-pass filter $\chi(k)$ to the Fourier Transform of r_θ to compute the envelope $a(t)$ and $b(t)$:

$$a(t) = 2\text{Re} \left\{ \mathcal{F}^{-1} [\widehat{r}_\theta(k+1)\chi(k)] \right\}, \quad (28)$$

$$b(t) = 2\text{Im} \left\{ \mathcal{F}^{-1} [\widehat{r}_\theta(k+1)\chi(k)] \right\}, \quad (29)$$

where $\chi(k)$ will be given later.

The low-pass filter $\chi(k)$ is determined by the choice of $V(\theta)$. The definition of $V(\theta)$ in (25) implies that $\chi(k)$ a stair function given as following:

$$\chi(k) = \begin{cases} 1, & -\lambda < k < \lambda \\ 0, & \text{otherwise.} \end{cases} \quad (30)$$

In the theoretical analysis in the subsequent section, we will see that the stair function is not a good choice as a filter. We can define a different space for $V(\theta)$ by choosing an appropriate $\chi(k)$. This opens up many choices for $V(\theta)$. In this paper, we choose the following low-pass filter $\chi(k)$ to define $V(\theta)$:

$$\chi(k) = \begin{cases} 1 + \cos \pi k, & -1 < k < 1 \\ 0, & \text{otherwise.} \end{cases} \quad (31)$$

Now, by incorporating the FFT-based solver in our iterative algorithm, we get the following FFT-based iterative algorithm:

- $\theta_k^0 = \theta_0$.

Step 1: Interpolate $r(t)$ to a uniform mesh in the θ^n -coordinate to get $r_{\theta^n}(\theta_j^n)$:

$$r_{\theta^n}(\theta_j^n) = \text{Interpolate } (\theta^n(t_i), r, \theta_j^n), \quad (32)$$

where θ_j^n , $j = 1, \dots, N$ are uniformly distributed in the θ^n -coordinate.

Step 2: Apply $\chi(k)$ to the Fourier Transform of r_{θ^n} to compute a^{n+1} and b^{n+1} on the mesh of the θ^n -coordinate:

$$a^{n+1}(\theta^n) = 2\text{Re} \{ \mathcal{F}^{-1} [\widehat{r}_{\theta^n}(k+1)\chi(k)] \}, \quad (33)$$

$$b^{n+1}(\theta^n) = 2\text{Im} \{ \mathcal{F}^{-1} [\widehat{r}_{\theta^n}(k+1)\chi(k)] \}. \quad (34)$$

Step 3: Interpolate $a^{n+1}(\theta^n)$ and $b^{n+1}(\theta^n)$ back to the uniform mesh of t :

$$a^{n+1}(t_i) = \text{Interpolate } (\theta_j^n, a^{n+1}(\theta_j^n), t_i), \quad (35)$$

$$b^{n+1}(t_i) = \text{Interpolate } (\theta_j^n, b^{n+1}(\theta_j^n), t_i). \quad (36)$$

Step 4: Update θ^n in the t -coordinate:

$$\theta_k^{n+1} = \theta_k^n - \lambda \arctan \left(\frac{b_k^{n+1}}{a_k^{n+1}} \right), \quad (37)$$

where $\lambda \in [0, 1]$ is chosen to make sure that θ_k^{n+1} is monotonically increasing:

$$\lambda = \max \left\{ \alpha \in [0, 1] : \frac{d}{dt} \left(\theta_k^n - \alpha \arctan \left(\frac{b_k^{n+1}}{a_k^{n+1}} \right) \right) \geq 0 \right\}. \quad (38)$$

Step 5: If $\|\theta_k^{n+1} - \theta_k^n\|_2 < \epsilon_0$, stop. Otherwise, go to Step 1.

In our implementation, instead updating the phase function θ , we update the instantaneous frequency $\theta'(t)$:

$$(\theta_k^{n+1})' = (\theta_k^n)' + \lambda \Delta \theta_k^{n+1}, \quad \Delta \theta_k^{n+1} = \frac{a_k^{n+1} (b_k^{n+1})' - b_k^{n+1} (a_k^{n+1})'}{(a_k^{n+1})^2 + (b_k^{n+1})^2}. \quad (39)$$

The phase function can be reconstructed from the instantaneous frequency by integration. The integration constant is set to 0 since this constant does not change the instantaneous frequency.

In the formula (39) to update θ , at the points where the denominator $(a_k^{n+1})^2 + (b_k^{n+1})^2$ is small, the error maybe amplified, then our algorithm may become unstable. In the real computation, we first set a threshold value α , at the points where the denominator is smaller than α , the value of $\Delta \theta$ is interpolated by the value $\Delta \theta$ at other points, its original value is not used. In the computations of this paper, α is set to be 0.1.

Above algorithm is based on the Gauss-Newton type iteration. It is known that this kind of iterations is sensitive to the initial value. It is not practical to assume that we can get a good initial guess especially when the signal is polluted by noise. In order to abate the dependence on the initial guess, we use the following procedure to gradually improve our approximation to the phase function so that it converges to the correct value.

First, for a given phase function θ , we define a set of spaces, such that

$$V_1(\theta) \subset V_2(\theta) \subset \cdots \subset V_K(\theta) = V(\theta) \quad (40)$$

In the computation, we first limit $\Delta \theta$ in corresponding $V_1(\theta)$ space, do the iteration until converge, then relax the restriction on $\Delta \theta$ to make it in $V_2(\theta)$, repeat the iteration until converge. Repeat this process until the algorithm converge with the restriction of $\Delta \theta$ in $V(\theta)$. In the real computations, this process converge even from very rough initial guess although we can not prove that.

This procedure is also very easy to implement. First, we apply a narrow low-pass filter on $\Delta \theta$, then make the filter broader and broader until it approach the filter given in (31).

In the above algorithm, we need to perform the Fourier Transform. In general, this works well only for periodic data. For a non-periodic signal with good scale separation property, we can extend the signal by a mirror reflection and treat the extended signal as a periodic signal. The result is still quite reasonable. But for nonperiodic data with poor scale separation property, the l^1 -regularized nonlinear matching pursuit is necessary to reduce the end effect, as we will see in Sections 4 and 5.

The initial guess of θ can be generated by other time-frequency analysis methods, such as the synchrosqueezed wavelet transforms [7]. In the following numerical examples, we obtain our initial guess using a simple approach based on the Fourier Transform. More precisely, by estimating the wavenumber by which the high frequency components are centered around, we can obtain a reasonably good initial guess for θ . The initial guess for θ obtained in this way is a linear function. As we will see in the following numerical examples, even with these

relatively rough initial guesses for θ , our algorithm still converges to the right answer with accuracy determined by the noise level.

Last point we want to remark is that our method has a close connection to the wavelet transform. In some sense, our method is equivalent to employ wavelet transform in the θ -coordinate in each step. The low-pass filter $\chi(k)$ used in our method can be related with the scale function in the multiresolution analysis. And space $V(\theta)$ can also be constructed following the idea of multiresolution analysis. The relation between our method and the wavelet transform will be studied more systematically in our future papers.

4 Numerical results

In this section, we will perform extensive numerical studies to demonstrate the effectiveness of our nonlinear matching pursuit method. First we will present numerical results for the FFT-based algorithms for periodic data or data with a good scale separation property (see section 6 for the definition of the scale separation property). In the second section, we will present numerical results for the l^1 regularized nonlinear matching pursuit which gives reasonably accurate decompositions for non-periodic data and even for under-sampled data or data with missing information in some physical domain.

4.1 Numerical results for the FFT-based algorithms

In this section, we will present a number of numerical experiments to demonstrate the accuracy and robustness of our method. We also compare the performance of our method with that of EMD or EEMD. A main focus of our numerical study is the robustness of the decomposition to signals that are polluted with a significant level of noise. When the signal is free of noise, we observe that the performance of our method is comparable to that of the EMD/EEMD method. However, when the signal is polluted by noise with a significant noise-to-signal ratio, our nonlinear matching pursuit method tends to give better performance than that of the EMD/EEMD method.

Throughout this section, we denote $X(t)$ as white noise with zero mean and variance $\sigma^2 = 1$. The Signal-to-Noise Ratio (SNR, measured in dB) is defined by

$$\text{SNR}[\text{dB}] = 10 \log_{10} \left(\frac{\text{var } f}{\sigma^2} \right). \quad (41)$$

We will apply our method to several different signals with increasing level of difficulty.

Example 1 The first example that we consider is a simple non-stationary signal consisting of a single IMF, which is given below

$$f(t) = \cos(60\pi t + 15 \sin(2\pi t)). \quad (42)$$

In Fig. 1, we plot the original signal on the left column and the instantaneous frequency on the right column. The curve with the red color corresponds to the exact instantaneous

frequency and the one with the blue color corresponds to the one obtained by our method. The top row corresponds to the original signal without noise, $f(t)$. The middle row corresponds to the same signal with a moderate noise level ($f(t)+X(t)$, SNR=-3.01dB) and the bottom row corresponds to the signal with large noise ($f(t)+3X(t)$, SNR=-12.55dB). In the case when no noise is added, the instantaneous frequency obtained by our method is almost indistinguishable from the exact one, see the first row of Fig. 1. When the signal has noise, our method can still extract the instantaneous frequency and corresponding IMF with reasonable accuracy, see Fig. 1 and Fig. 2.

In Fig. 2, we compare the IMFs extracted by our method with those obtained by the EMD/EEMD method. For the signal without noise, we use the EMD method to decompose the signal. For the signal with noise, we use the EEMD method to decompose the signal. In the EEMD approach, the number of ensembles is chosen to be 200 and the standard deviation of the added noise is 0.2. In each ensemble, the number of sifting is set to 8. Even though the signal we consider has only a single IMF, the EEMD method still produces several IMFs. Among different components of IMFs that are produced by the EEMD method, we select the one that is closest to the exact IMF in l^2 norm and display it in Fig. 2.

When the signal does not have noise, both our method and the EMD method produce qualitatively the same result for this simple signal, see the first row of Fig. 2. When noise is added, the situation is quite different. The IMFs extracted by our method still have reasonable accuracy. However, the IMF decomposed by EEMD fails to capture the phase of the exact IMF in some region. As a consequence, the accuracy of the instantaneous frequency obtained by the EEMD method is very poor (not shown here).

Example 2

Now, we consider a signal that consists of three IMFs.

$$f(t) = \frac{1}{1.5 + \cos(2\pi t)} \cos(60\pi t + 15 \sin(2\pi t)) + \frac{1}{1.5 + \sin(2\pi t)} \cos(160\pi t + \sin(16\pi t)) + (2 + \cos(8\pi t)) \cos(140\pi(t+1)^2). \quad (43)$$

In Fig. 3, we study the accuracy of the instantaneous frequencies obtained by our method with the exact instantaneous frequencies. The upper row corresponds to the signal without noise. As we can see, it is hard to tell any hidden structure from this signal even without noise. Our method recovers the three components of the instantaneous frequencies (blue) that match the exact instantaneous frequencies (red) extremely well. They are almost indistinguishable from each other. In the case when noise is added to the original signal, the polluted signal looks really complicated and one can not recognize any hidden pattern from this polluted signal. It is quite amazing that our method could still recover the three components of the instantaneous frequencies with accuracy comparable to the noise level, see the bottom row of Fig. 3.

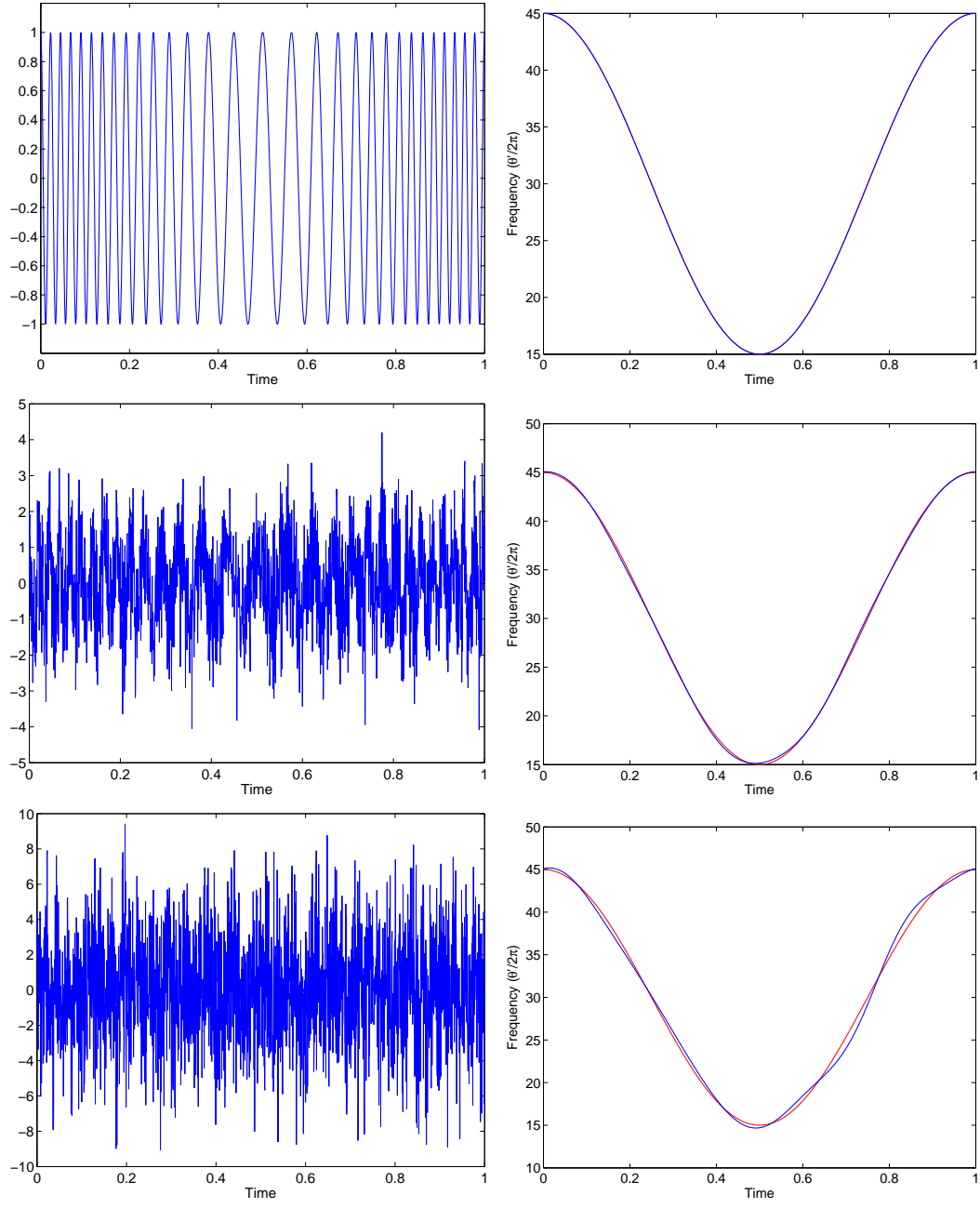


Figure 1: Top row: left: the original signal defined by (42) without noise; right: Instantaneous frequencies; red: exact frequency; blue: numerical results. Middle row: the same as the top row except white noise $X(t)$ is added to the original signal, the corresponding SNR is -3.01 dB. Bottom row: White noise $3X(t)$ is added to the original signal, the corresponding SNR is -12.55 dB.

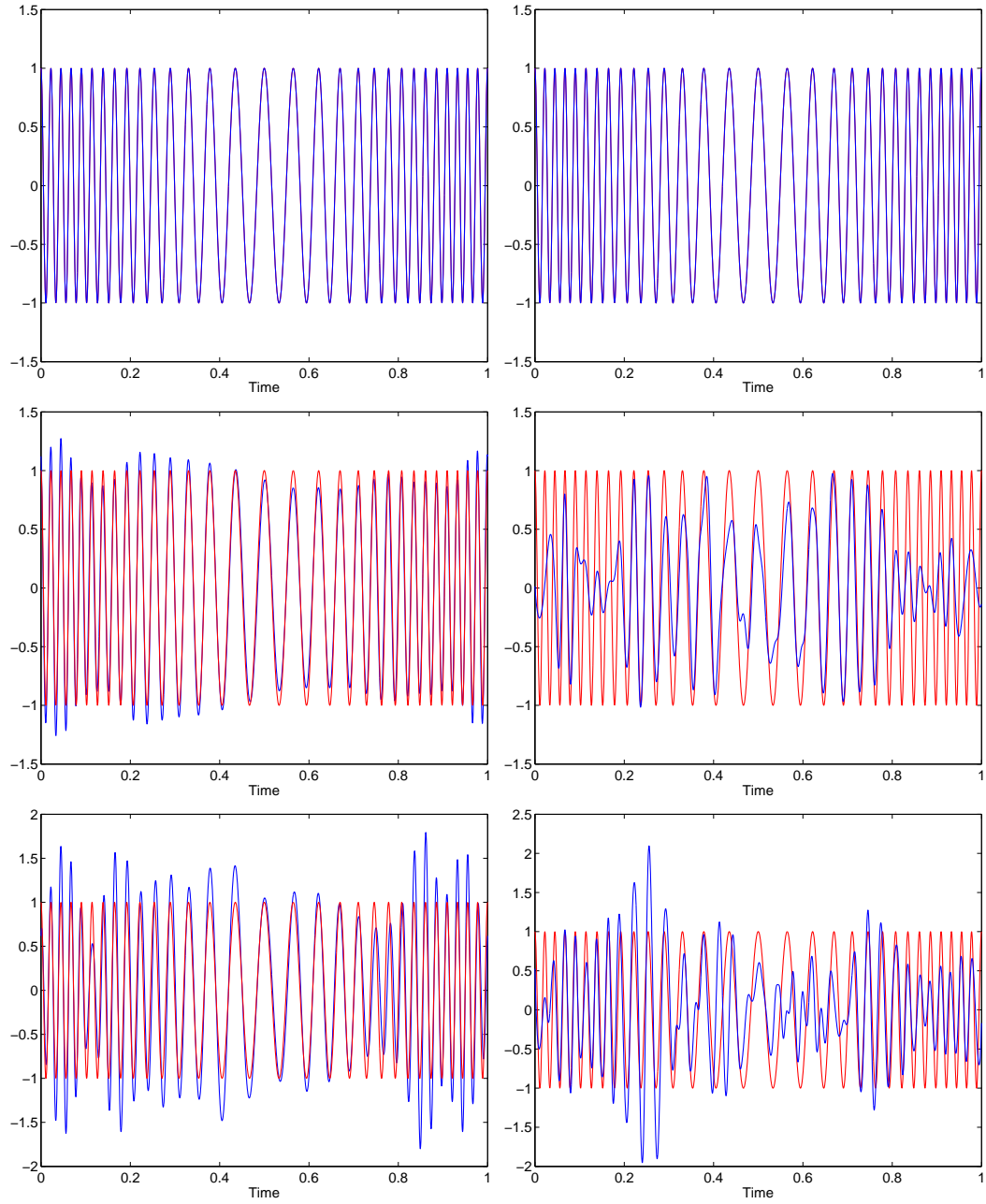


Figure 2: The IMFs extracted by our method and EMD/EEMD method. Left column: IMFs extracted by our method; Right column: IMFs obtained by EMD/EEMD method. Top row: IMFs from $f(t)$; Middle row: IMFs from $f(t) + X(t)$; Bottom row: IMFs from $f(t) + 3X(t)$. $f(t)$ is defined in (42).

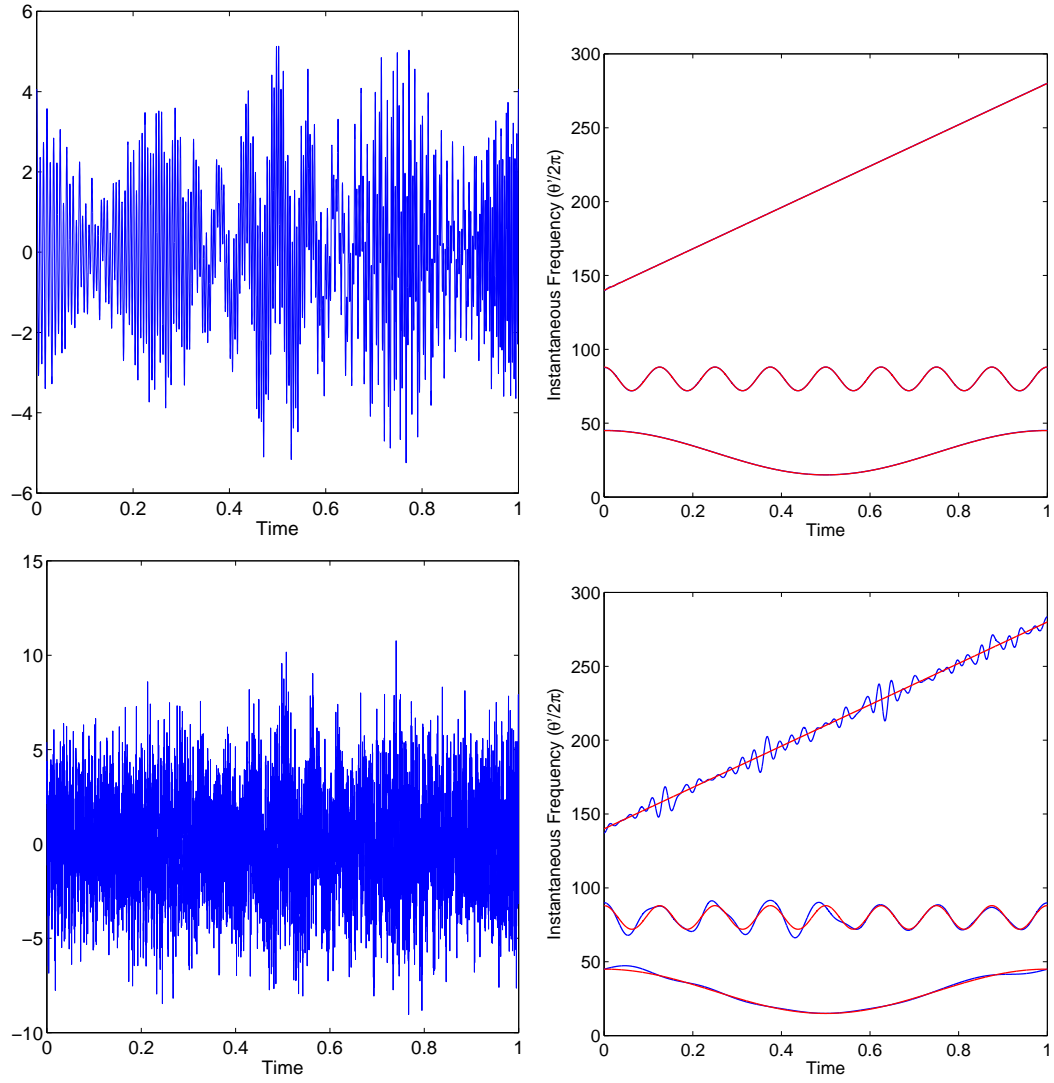


Figure 3: Upper row: left: the signal defined in (43) without noise; right: Instantaneous frequencies; red: exact frequencies; blue: numerical results. Lower row: the same as the upper row except that white noise $2X(t)$ was added to the original signal, the corresponding SNR is -0.8 dB.

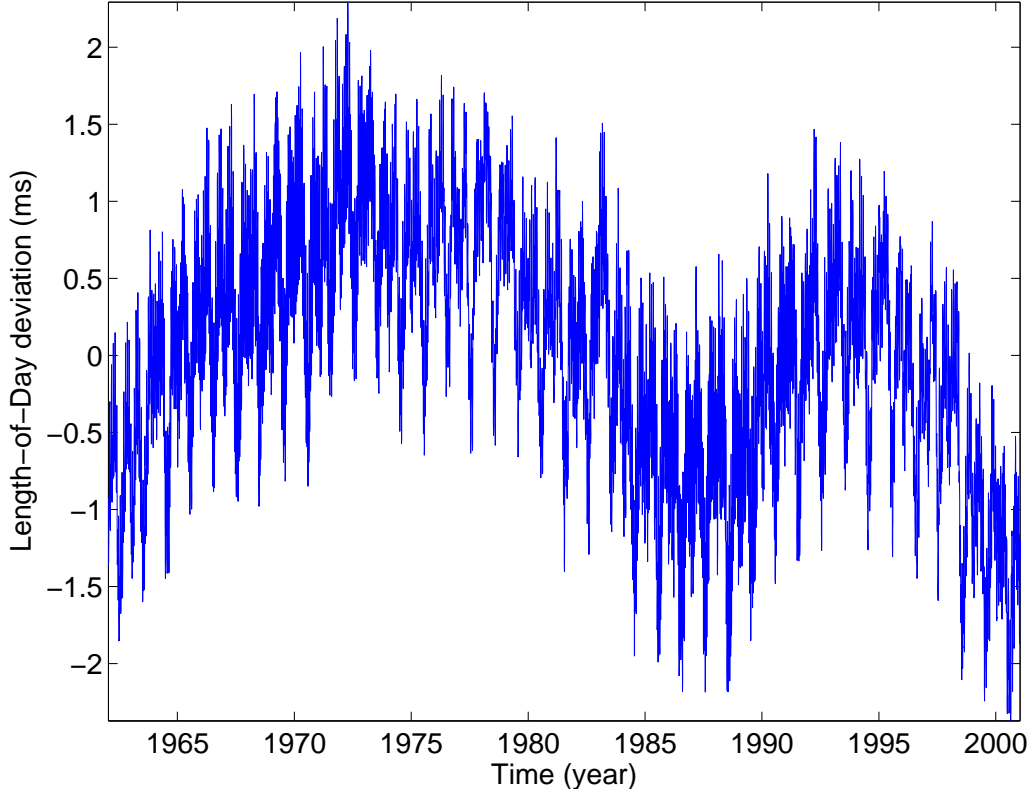


Figure 4: The daily Length-of-Day data from Jan 20, 1962 to Jan 6, 2001.

Real Data: Length-of-Day Data

Next, we apply our method to the Length-of-Day data, see Fig. 4. The data we adopt here was produced by Gross [10], covering the period from 20 January 1962 to 6 January 2001, for a total of 14,232 days (approximate 39 years). In our previous paper [13], we also studied this data set. Due to the high computational cost associated with the l^1 minimization, we can not decompose the entire data set. Instead, we decompose a segment of the data that contains 700 consecutive days. Thanks to the low computational cost of the FFT-based nonlinear matching pursuit method, we can now study the entire data set without any compromise.

Fig. 5 displays the first 5 IMFs extracted by the FFT-based method. These IMFs are sorted by their frequencies from high to low. We note that the results obtained by our method do not suffer from the mode mixing phenomenon that is present in the EMD decomposition. Moreover, each component is enforced to be an IMF by the construction of our dictionary. Thus, there is no need to do shifting or post-processing as was done in the EMD or the EEMD method. And the IMFs we got are qualitatively match those obtained by EEMD with post-processing [30].

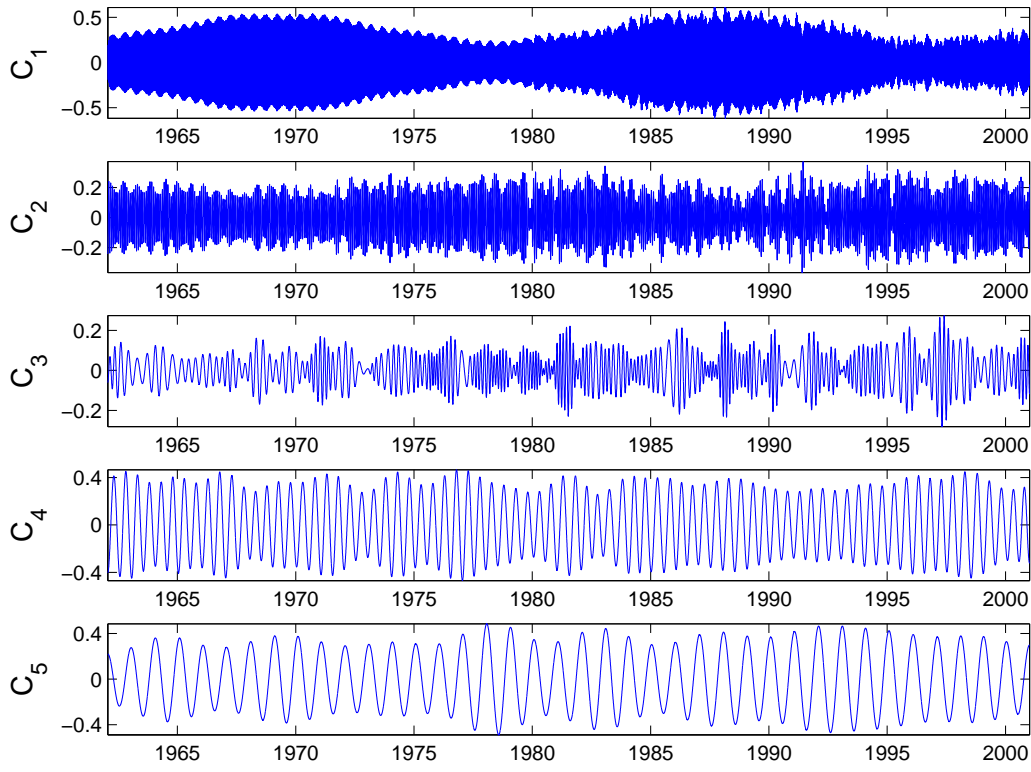


Figure 5: The first 5 IMFs with highest frequencies given by our FFT-based method.

It is interesting to note that each IMF that we obtain has a clear physical interpretation. For example, the period of C_1 is around 14 days, corresponding to the semi-monthly tides. The period of C_2 is about 28 days, corresponding to the monthly tides. Similarly, the period of C_4 is about half a year, corresponding to the semi-annual cycle and C_5 corresponds to the annual cycle.

4.2 Numerical results for the l^1 regularized nonlinear matching pursuit

The nonlinear matching pursuit method based on the Fourier Transform works well only for periodic data. For non-periodic data or data with poor scale separation, the results obtained by this method tend to produce some oscillations near the boundary. This so-called “end effect” is also present in the EMD method and other data analysis methods. In our method, the “end effect” comes from the use of the Fourier Transform in the algorithm. To remove this “end effect” error, we need to use the l^1 regularized nonlinear matching pursuit described in Section 3.2 with $V(\theta)$ being the overcomplete Fourier basis defined in (2).

4.2.1 Numerical results for non-periodic data

In this subsection, we perform a numerical experiment to test the effectiveness of our l^1 regularized nonlinear matching pursuit for non-periodic data. We first consider the following data in our experiment:

$$\begin{aligned} \theta_1 &= 20\pi(t+1)^2 + 1, \quad \theta_2 = 161.4\pi t + 4(1-t)^2 \sin(16\pi t), \\ f(t) &= \frac{1}{1.5 + \sin(1.5\pi t)} + (2t+1) \cos \theta_1 + (2-t)^2 \cos \theta_2. \end{aligned} \quad (44)$$

In this numerical example, the parameter γ is chosen to be 1. From Fig. 6, we observe that the l^1 regularized nonlinear matching pursuit seems to produce considerably smaller error near the boundary for this non-periodic signal.

4.2.2 Numerical results for data with incomplete or sparse samples

The l_1 regularized nonlinear matching pursuit can also handle the incomplete data and the data with sparse samples. We illustrate this property of our method through a few examples.

The first example is an incomplete signal given by (45).

$$\theta(t) = 120\pi t + 10 \cos(4\pi t), \quad a(t) = 2 + \cos(2\pi t), \quad f(t) = a(t) \cos \theta(t), \quad t \in [0, 0.4] \cup [0.6, 1]. \quad (45)$$

For this signal, we have only eighty percent of the original data and miss twenty percent of the data in the gap interval $[0.4, 0.6]$. In Fig. 7, we plot the recovered signal in the gap interval $[0.4, 0.6]$ (see the middle panel). The recovered signal matches the original signal almost perfectly in the gap interval. The recovered instantaneous frequency also matches the exact instantaneous frequency with high accuracy.

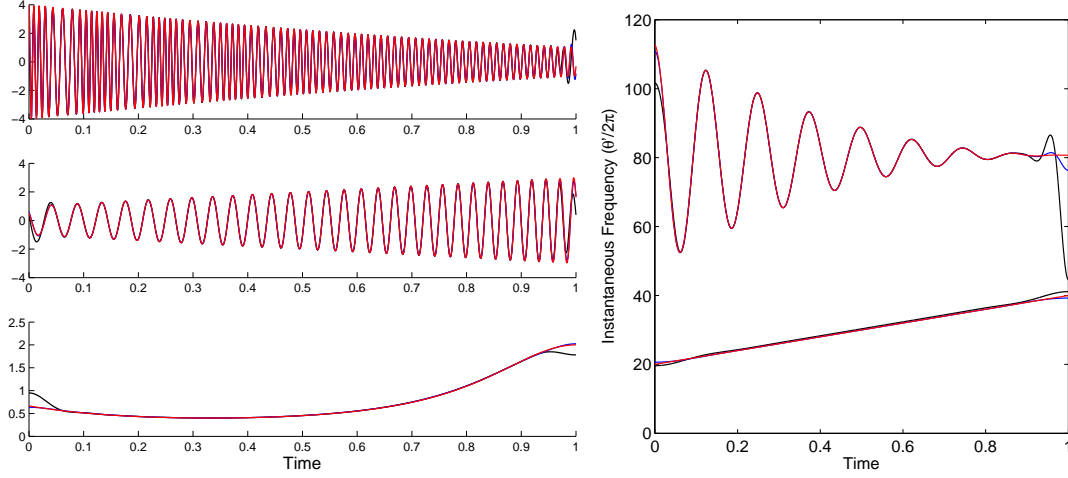


Figure 6: IMF (left) and Instantaneous frequency (right) of the signal in (44) obtained from different methods. Red: exact; Blue: l^1 regularized nonlinear matching pursuit; Black: FFT-based algorithm.

In Fig. 8, we perform the same numerical experiment by enlarging the interval of missing data from $(0.4, 0.6)$ to $(0.3, 0.7)$, i.e. we miss forty percent of the data. Even for this more challenging example, our method still gives quite reasonable reconstruction of the original data in the region of missing data. The recovered instantaneous frequency still approximates the exact instantaneous frequency with reasonable accuracy, especially away from the region of missing data.

Finally, we consider an example with insufficient samples. The signal is generated as follows:

$$\theta(t_i) = 120\pi t_i + 10 \cos(2\pi t_i), \quad a(t_i) = 2 + \cos(2\pi t_i), \quad f(t_i) = a(t_i) \cos \theta(t_i), \quad t_i \in [0, 1]. \quad (46)$$

and $i = 1, 2, \dots, N$. The location t_i is chosen randomly in $[0, 1]$. In this example, the number of samples is 64. This means that we have about one sample point within one period of the signal on average.

Fig. 9 gives the results obtained by our method. In the case of insufficient samples without noise, the recovered signal and the original signal are almost indistinguishable.

We now add Gaussian noise to the original samples and apply our method to this noisy data. The result is given in Fig. 10. In this case, the noise $0.2X(t)$ is added to the original signal $f(t)$ given in (46). We can see that both the recovered signal and the instantaneous frequency still have reasonable accuracy. This shows that our method is stable with respect to noise perturbation even for sparse under-sampled data.

In the future, we plan to perform some theoretical study of our method for data with incomplete or sparse samples and carry out more numerical experiments for some complicated or real data.

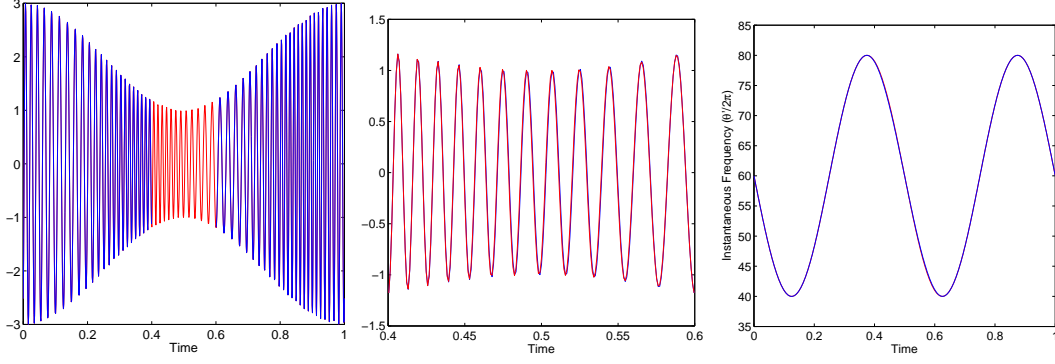


Figure 7: Left: blue: the original incomplete data, the gap is $(0.4, 0.6)$; red: the missing data recovered by our method; Middle: the recovered missing data, red: exact; blue: numerical. Right: the instantaneous frequencies, red: exact; blue: numerical.

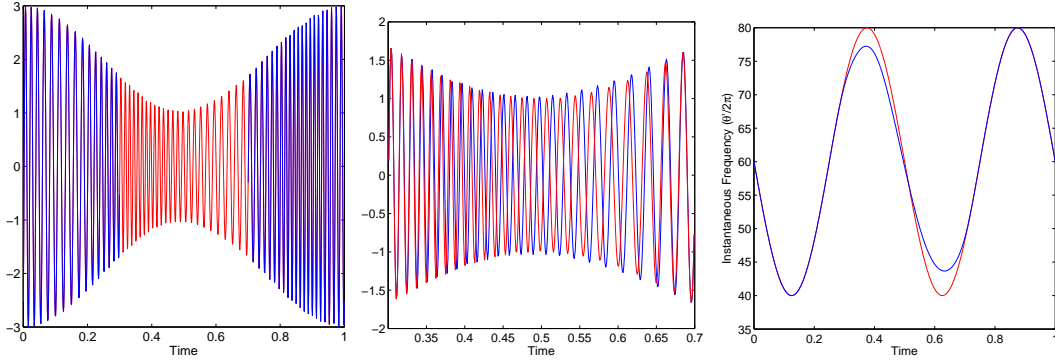


Figure 8: Left: blue: the original incomplete data, the gap is $(0.3, 0.7)$; red: the missing data recovered by our method; Middle: the recovered missing data, red: exact; blue: numerical. Right: the instantaneous frequencies, red: exact; blue: numerical.

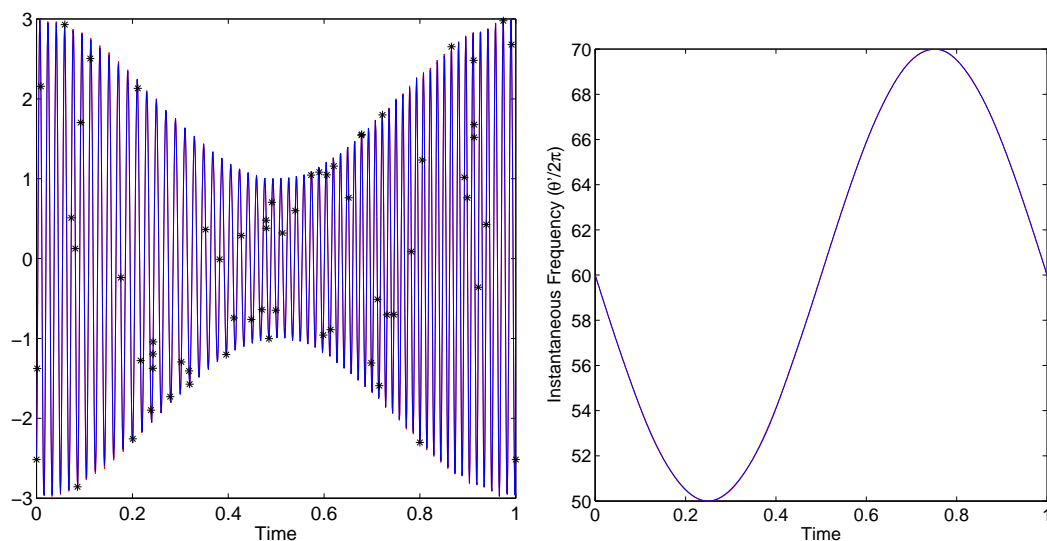


Figure 9: Left: original samples, red: exact; blue: recovered; '*' represent the sample points. Right: instantaneous frequency, red: exact; blue: numerical.

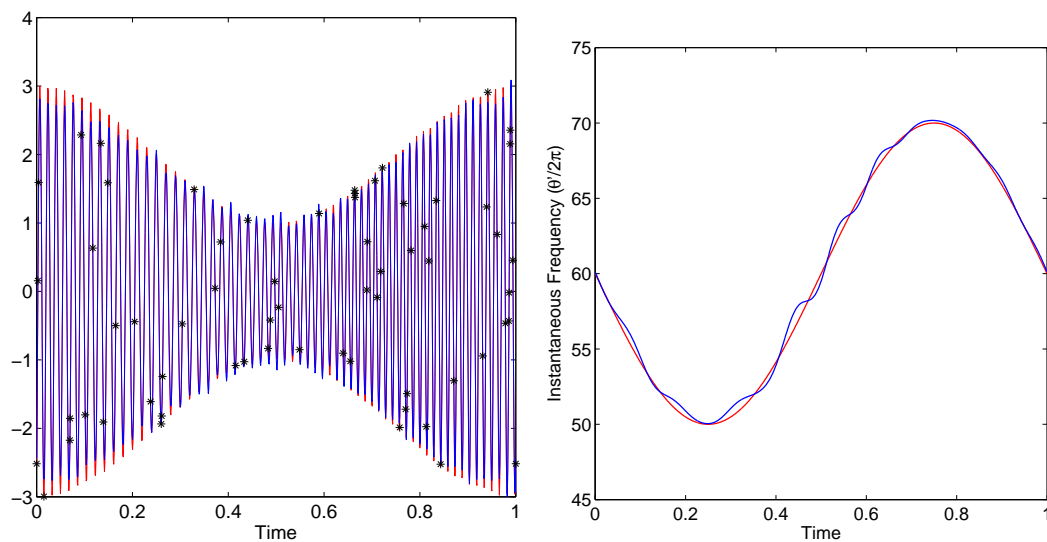


Figure 10: Left: original samples, red: exact; blue: recovered from the noised data, $f(t_i) + 0.2X(t_i)$; '*' represent the sample points. Right: instantaneous frequencies, red: exact; blue: numerical.

5 Generalizations for the l^1 regularized nonlinear matching pursuit

The iterative algorithm based on l^1 regularized nonlinear matching pursuit can also be generalized to deal with more complicated data, such as the data with poor scale-separation and the data with intra-wave frequency modulation. In this section, the power of this kind of algorithm is shown through several numerical examples, more details can be presented in our subsequent papers.

5.1 Numerical results for data with poor scale-separation property

In the previous sections, we show that for data with a good scale-separation property, the algorithm based on l^1 regularized nonlinear matching pursuit can give an accurate decomposition. Now, we give a brief discussion how to decompose data with poor scale-separation property.

Let f be a signal that has the following sparse decomposition over dictionary \mathcal{D} :

$$f(t) = \sum_{k=1}^M a_k \cos \theta_k, \quad a_k \cos \theta_k \in \mathcal{D}, \quad (47)$$

where \mathcal{D} is defined in (17). But now the instantaneous frequencies $\theta'_k(t)$ are not well separated, so $f(t)$ does not satisfy the scale-separation condition defined in Section 6.

It is well known that for data consisting of components with interfering frequencies, the matching pursuit with a Gabor dictionary may not give a sparse decomposition [19]. Since our method is based on the matching pursuit, it is not surprising that it may not be able to generate the sparsest decomposition either.

To illustrate, we consider the following signal consisting of two IMFs whose instantaneous frequencies intersect each other. The signal is generated by the analytical formula given below.

$$f(t) = \cos(20\pi t + 40\pi t^2 + \sin(2\pi t)) + \cos(40\pi t). \quad (48)$$

Fig. 11 plots the instantaneous frequencies and IMFs recovered by the nonlinear matching pursuit given in the previous section. Near the point of intersection, both the instantaneous frequencies and IMFs produce noticeable errors. The good news is that the instantaneous frequency recovered by our method is still in phase with the exact one. Furthermore, the accuracy is quite reasonable in the region far away from the point of intersection. This shows that our method has a temporal locality property, which is important in many physical applications.

To further improve the accuracy of our decomposition when there are a number of instantaneous frequencies that are not well separated, we need to decompose these components simultaneously since these IMFs have strong correlation. Assume that we can learn from our l^1 regularized nonlinear matching pursuit that there are M components of IMFs whose

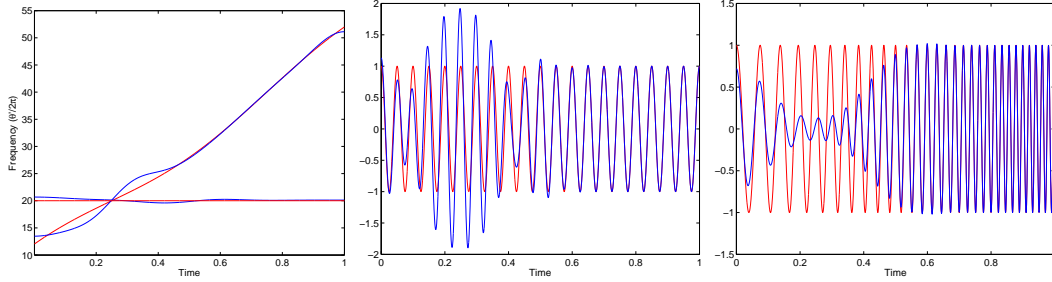


Figure 11: Left: Instantaneous frequencies; red: exact frequencies; blue: numerical results. Middle and Right: IMFs extracted by the previous nonlinear matching pursuit.

instantaneous frequencies are not well separated, we modify our decomposition method to solve the following optimization problem:

$$\min \left(\gamma \sum_{k=1}^M \|\hat{a}_k\|_{l^1} + \|f(t) - \sum_{k=1}^M a_k \cos \theta_k\|_{l^2}^2 \right), \quad s.t. \quad a_k \cos \theta_k \in \mathcal{D}, \quad (49)$$

where $\gamma > 0$ is a regularized parameter and \hat{a}_k is the representation of a_k in $V(\theta_k)$ space.

This problem is much more difficult to solve than the original one, since the different components may have strong correlation. Based on the l^1 regularized nonlinear matching pursuit that we introduced in the previous sections, we have developed a new method to solve the above optimization problem. The detail of this method will be reported in our subsequent paper. Here we give an example to demonstrate that this new method has the capability to deal with the signal with poor scale separation.

Fig 12 gives the results obtained by our new method for the signal given in (48). We can see that both the instantaneous frequencies and IMFs match the exact ones pretty well. These results are much better than those given in Fig. 11.

In the previous example we consider, the scale separation is destroyed near the region where instantaneous frequencies of two different IMFs intersect each other. In the following example, we consider another example with poor scale separation, but for a different reason. In this case, the frequencies of different IMFs are well separated, but the instantaneous frequency of an IMF is not smooth any more. Specifically, we consider the following example:

$$f = \frac{1}{1.5 + \cos 2\pi t} + \cos \theta_1 + \cos \theta_2, \quad \theta_1(0) = \theta_2(0) = 0, \quad (50)$$

$$\theta'_1 = \begin{cases} 40\pi, & 0 \leq t \leq 0.3, \\ 60\pi, & 0.3 < t \leq 1. \end{cases}, \quad \theta'_2 = \begin{cases} 140\pi, & 0 \leq t \leq 0.6, \\ 160\pi, & 0.6 < t \leq 1. \end{cases}.$$

In this example, the instantaneous frequencies are discontinuous at $t = 0.3$ and $t = 0.6$ respectively. Even for such signal, our l^1 regularized nonlinear matching pursuit method still gives a reasonable decomposition even if the signal is polluted with noise.

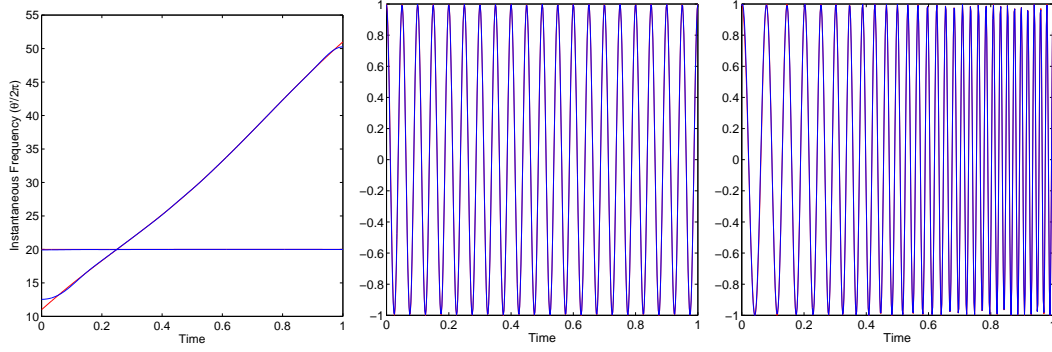


Figure 12: Left: Instantaneous frequencies; Middle (first component) and right (second component): IMFs obtained by extracting two IMFs together. red: exact results; blue: numerical results.

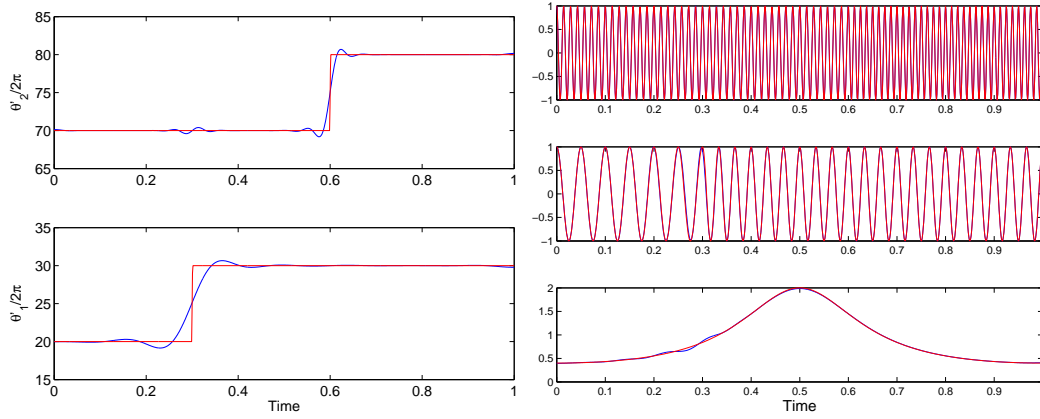


Figure 13: Instantaneous frequencies and IMFs given by the l^1 regularized nonlinear matching pursuit for the data given in (50). Left: Instantaneous frequencies; Right: IMFs obtained by our method. red: exact results; blue: numerical results.

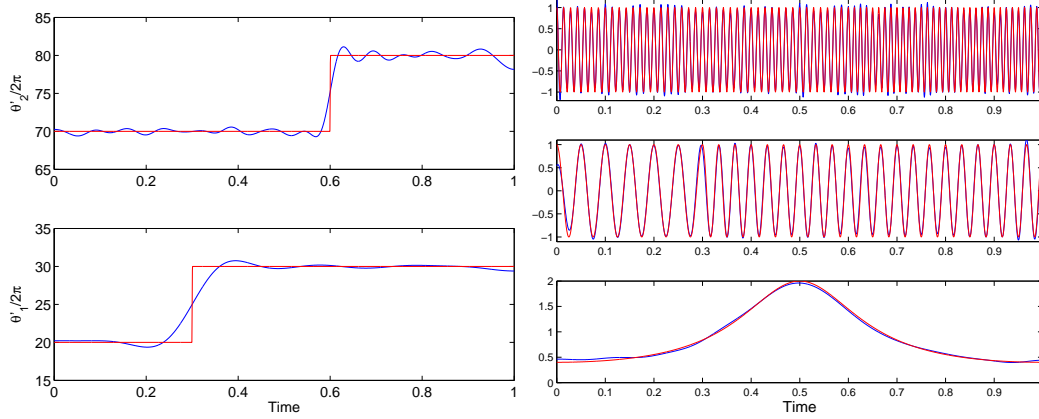


Figure 14: Instantaneous frequencies and IMFs given by l^1 regularized nonlinear matching pursuit for the data $f(t) + 0.3X(t)$, where f is given in (50) and $X(t)$ is Gaussian noise with standard derivation $\sigma^2 = 1$. Left: Instantaneous frequencies; Right: IMFs obtained by our method. red: exact results; blue: numerical results.

As we can see from Fig. 13, there are some oscillations due to the Gibbs phenomenon near the points of discontinuity ($t = 0.3, 0.6$). IMFs also have some errors near these two points. But in the region away from these two points, the numerical results match very well with the exact ones. These results also suggest that our l^1 regularized nonlinear matching pursuit has temporal locality property. The error is confined in a small region near the points where scale separation is poor. This property will be discussed further in Section 6. When the noise is added to this signal, the results we obtain still have a reasonable accuracy, see Fig. 14.

5.2 Numerical results for data with intra-wave frequency modulation

In some physical applications such as the stokes waves or some nonlinear dynamic systems, we need to deal with data that have intra-wave frequency modulation. For this type of signals, they have the sparse decomposition:

$$f(t) = \sum_{k=1}^M a_k \cos \theta_k, \quad (51)$$

where a_k are smooth envelopes, but their instantaneous frequencies, θ'_k , are not smooth any more. Typically, the phase function has the form $\theta_k = \phi_k + \epsilon \cos(\omega_k \phi_k)$, where ϕ_k is a smooth function, $\epsilon > 0$ is a small number and $\omega_k \geq 1$ is an integer. In the study of some nonlinear waves or dynamical systems, ω_k is an important parameter related to the characteristic of the nonlinearity of the system.

An essential difficulty for this type of data is that the instantaneous frequency, θ'_k , is as oscillatory as $\cos \theta_k$ or even more oscillatory. In the nonlinear matching pursuit method

that we proposed in the sections, we assume that a_k and b_k are smoother than $\cos \theta_k$ and use these two coefficients to update θ_k . In the case where the signal has strong intra-wave modulation, θ'_k is as oscillatory as $\cos \theta_k$. Thus our current method would not be able to give a good approximation of θ'_k unless we are given a very good initial guess for θ_k . To overcome this difficulty, we introduce a shape function, s_k , to replace cosine function. The idea is to absorb the high frequency intra-wave modulation into the shape function s_k . This will ensure that θ'_k is smoother than $s_k(\theta_k)$. This idea has been proposed by Hau-tieng Wu in [29], but efficient algorithm to compute the shape function has not appear in the literature.

Note that s_k is not known *a priori* and is adapted to the signal. We need to learn s_k from the physical signal. This consideration naturally motivates us to modify the construction of the dictionary as follows:

$$\mathcal{M} = \{a_k s_k(\theta_k) : a_k, \theta'_k \in V(\theta_k), s_k \text{ is } 2\pi\text{-period function}\}, \quad (52)$$

where s_k is an unknown 2π -periodic ‘shape function’ and is adapted to the signal. If we choose s_k to be cosine function, then $\mathcal{M} = \mathcal{D}$.

We have developed an efficient method based on our l^1 regularized nonlinear matching pursuit to recover the components with intra-wave frequency modulation by looking for the sparsest decomposition using this new dictionary \mathcal{M} . By exploring the fact that s_k is a periodic function of θ_k , we can identify certain low rank structure of the signal. This structure enables us to extract the shape function from the signal. Once we get an approximation of the shape function s_k , we can use s_k to update θ_k . This process continues until it converges. The detail of this method will appear in another paper. Here, we give several numerical examples to demonstrate the capability of our method.

The first example is the solution of the Duffing equation. This is an important example to demonstrate the importance of the intra-wave frequency modulation. Our method can handle this signal even with noise.

The Duffing equation is a nonlinear ODE which has the following form:

$$\frac{d^2 u}{dt^2} + u + \epsilon u^{1+\omega} = \gamma \cos(\beta t). \quad (53)$$

The parameters, ϵ, γ, ω , that we use here to generate the solution in Fig. 15, are the same as those in the paper [14], $\epsilon = -1$, $\gamma = 0.1$, $\beta = \frac{1}{25}$ and $\omega = 2$. The initial condition is $u(0) = u'(0) = 1$.

In Fig. 15, we plot the shape function that we obtain from the solution of the Duffing equation. In this example, we can express $s_k(\theta_k)$ in terms of $\cos \tilde{\theta}_k$, from which we can recover the instantaneous frequency of the signal. More interestingly, from the Fourier coefficients of the shape function, we can recover the information regarding the nonlinearity of the Duffing equation, which is $\omega = 2$ in this case. We refer to a recent work of Prof. Norden Huang for more discussions of how to extract the degree of nonlinearity using EMD (see Dr. Huang’s lecture in the IMA Hot Topic Workshop on Trend and Instantaneous Frequency, September 7-9, 2011, IMA).

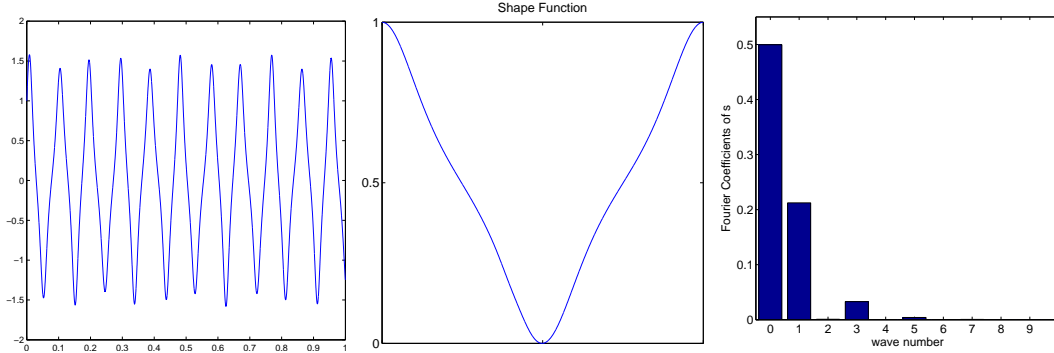


Figure 15: Left: the solution of duffing equation; Middle: the shape function s ; Right: the Fourier coefficients of s .

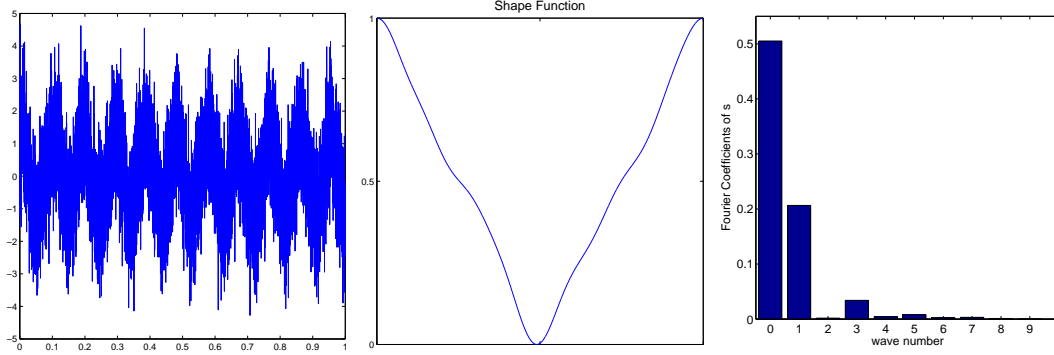


Figure 16: Left: the solution of duffing equation with noise $X(t)$; Middle: the shape function s ; Right: the Fourier coefficients of s .

We also add Gaussian noise $X(t)$ with covariance $\sigma^2 = 1$ to the original solution of the Duffing equation. Fig. 16 shows the corresponding results. We can see that the shape function extracted from the noisy signal still keeps the main characteristics of the shape function extracted from the signal without noise. We can also clearly extract the degree of nonlinearity, $\omega = 2$, even with such large noise perturbation to the solution of the Duffing equation.

6 Some preliminary error analysis for the data with scale separation

In this section, we perform some preliminary error analysis for our nonlinear matching pursuit method. To guarantee uniqueness of the decomposition, we need to impose certain scale separation property for the data that we try to decompose. Before we state our result,

we first define what we mean by scale separation for a given signal.

Definition 6.1 (Scale-separation) *One function $f(t) = a(t) \cos \theta(t)$ is said to satisfy a scale-separation property with a separation factor $\epsilon > 0$, if $a(t)$ and $\theta(t)$ satisfy the following conditions:*

$$\begin{aligned} a(t) &\in C^1(\mathbb{R}), \quad \theta \in C^2(\mathbb{R}), \\ \inf_{t \in \mathbb{R}} \theta'(t) &> 0, \quad M = \sup_{t \in \mathbb{R}} |\theta''(t)| < \infty \\ \left| \frac{a'(t)}{\theta'(t)} \right|, \quad \left| \frac{\theta''(t)}{(\theta'(t))^2} \right| &\leq \epsilon, \quad \forall t \in \mathbb{R}. \end{aligned}$$

Definition 6.2 (Well-separated signal) *A signal $f : \mathbb{R} \rightarrow \mathbb{R}$ is said to be well-separated with separation factor ϵ and frequency ratio $d > 1$ if it can be written as*

$$f(t) = \sum_{k=1}^K a_k(t) \cos \theta_k(t)$$

where all $f_k(t) = a_k(t) \cos \theta_k(t)$ satisfies the scale-separation property with separation factor ϵ , and their phase function θ_k satisfies

$$\theta'_k(t) \geq d\theta'_{k-1}(t), \quad \forall t \in \mathbb{R}. \quad (54)$$

Theorem 6.1 *Let $f(t)$ be a function satisfying the scale-separation property with separation factor ϵ and frequency ratio d as defined in Definition 6.2. Choose a low-pass filter ϕ such that its Fourier Transform $\hat{\phi}$ has support in $[-\Delta, \Delta]$ with $\Delta < \frac{d-1}{d+1/2}$ and $\hat{\phi}(k) = 1, \forall k \in [-\Delta/2, \Delta/2]$. If in the n th step, the approximate phase function $\theta_{k_0}^n$ satisfies the following condition:*

$$\left| \frac{(\theta_{k_0}^n)'(t)}{\theta_{k_0}'(t)} - 1 \right| < \frac{\Delta}{2}, \quad \left| \frac{(\theta_{k_0}^n)''(t)}{\left((\theta_{k_0}^n)'(t)\right)^2} \right| \leq \epsilon, \quad (55)$$

then the accuracy of the phase function in the next step is order ϵ , i.e.

$$\left| \theta_{k_0}^{n+1}(t) - \theta_{k_0}(t) \right| = O(\epsilon). \quad (56)$$

In order to prove the above theorem, we need the following lemma:

Lemma 6.1 *For any $a(t) \in C^1(\mathbb{R}), \theta \in C^2(\mathbb{R})$, we have*

$$\left| \int a(\tau) e^{-i\theta(\tau)} \phi(\tau - t) d\tau - a(t) e^{-i\theta(t)} \hat{\phi}(\theta'(t)) \right| \leq \sup |a'(t)| I_1 + \frac{1}{2} |a(t)| \sup |\theta''(t)| I_2, \quad (57)$$

where $I_n = \int |t^n \phi(t)| dt$.

Proof The proof follows from the following direct calculations:

$$\begin{aligned}
& \left| \int a(\tau) e^{-i\theta(\tau)} \phi(\tau - t) d\tau - a(t) e^{-i\theta(t)} \widehat{\phi}(\theta'(t)) \right| \\
&= \left| \int (a(\tau) - a(t)) e^{i\theta(\tau)} \phi(\tau - t) d\tau + a(t) \int \left(e^{-i\theta(\tau)} - e^{-i(\theta(t) - \theta'(t)(\tau - t))} \right) \phi(\tau - t) d\tau \right| \\
&= \left| \int (a(\tau) - a(t)) e^{i\theta(\tau)} \phi(\tau - t) d\tau + a(t) \int \left(e^{-i(\theta(\tau) - \theta(t) - \theta'(t)(\tau - t))} - 1 \right) e^{-i(\theta(t) - \theta'(t)(\tau - t))} \phi(\tau - t) d\tau \right| \\
&\leq \sup |a'(t)| \int |\tau \phi(\tau)| d\tau + |a(t)| \left| \int \left(e^{-\frac{1}{2}i\theta''(s(\tau))(\tau - t)^2} - 1 \right) e^{-i(\theta(t) - \theta'(t)(\tau - t))} \phi(\tau - t) d\tau \right| \\
&\leq \sup |a'(t)| \int |\tau \phi(\tau)| d\tau + |a(t)| \int \left| \frac{1}{2} \theta''(s(\tau)) (\tau - t)^2 \phi(\tau - t) \right| d\tau \\
&\leq \sup |a'(t)| \int |\tau \phi(\tau)| d\tau + \frac{1}{2} |a(t)| \sup |\theta''(t)| \int |\tau^2 \phi(\tau)| d\tau \\
&= \sup |a'(t)| I_1 + \frac{1}{2} |a(t)| \sup |\theta''(t)| I_2
\end{aligned} \tag{58}$$

□

Remark 6.1 We remark that since we typically deal with data of finite support and extend them periodically to the whole domain, the estimates for I_1 and I_2 in the above lemma are effectively taken only in the finite support of the data.

Corollary 6.1 If the Fourier Transform of the low-pass filter ϕ is symmetric, i.e. $\widehat{\phi}(k) = \widehat{\phi}(-k)$, then we have

$$\left| \int a(\tau) \cos(\theta(\tau)) \phi(\tau - t) d\tau - a(t) \cos \theta(t) \widehat{\phi}(\theta'(t)) \right| \leq \sup |a'(t)| I_1 + \frac{1}{2} |a(t)| \sup |\theta''(t)| I_2 \tag{59}$$

$$\left| \int a(\tau) \sin(\theta(\tau)) \phi(\tau - t) d\tau - a(t) \sin \theta(t) \widehat{\phi}(\theta'(t)) \right| \leq \sup |a'(t)| I_1 + \frac{1}{2} |a(t)| \sup |\theta''(t)| I_2. \tag{60}$$

Now we can prove Theorem 6.1. Here we only give a sketch of the proof, the detail of the proof is deferred to the Appendix.

Proof of Theorem 6.1: In order to simplify the notation, we denote $\bar{\theta} = \theta_{k_0}^n$ and define $\overline{(\cdot)}$ as the mapping from t to $\bar{\theta}$, i.e. $\overline{f}(\bar{\theta}) = f(t)$, $\forall f$. In our algorithm, we update $\theta_{k_0}^{n+1}$ by the following step

$$\theta_{k_0}^{n+1} = \bar{\theta} - \arctan \left(\frac{b(t)}{a(t)} \right), \quad a(t) = A(\bar{\theta}(t)), \quad b(t) = B(\bar{\theta}(t)), \tag{61}$$

where

$$A(\gamma) = 2 \int \overline{f}(\bar{\theta}) \cos(\bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta}, \quad B(\gamma) = 2 \int \overline{f}(\bar{\theta}) \sin(\bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta}. \tag{62}$$

We will prove that $a(t)$ and $b(t)$ satisfy the following estimates:

$$a(t) = A(\bar{\theta}(t)) = a_{k_0}(t) \cos(\theta_{k_0}(t) - \bar{\theta}) + O(\epsilon). \quad (63)$$

$$b(t) = B(\bar{\theta}(t)) = a_{k_0}(t) \sin(\theta_{k_0}(t) - \bar{\theta}) + O(\epsilon). \quad (64)$$

Then, we can get

$$\Delta\theta = \arctan\left(\frac{B(t)}{A(t)}\right) = \theta_{k_0}(t) - \bar{\theta} + O(\epsilon), \quad (65)$$

which implies that

$$\left|\theta_{k_0}^{n+1}(t) - \theta_{k_0}(t)\right| = O(\epsilon). \quad (66)$$

First, let us estimate $A(\gamma)$ as follows:

$$\begin{aligned} A(\gamma) &= 2 \int \bar{f}(\bar{\theta}) \cos(\bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta} = 2 \sum_{k=1}^n \int \bar{a}_k(\bar{\theta}) \cos \theta_k(t) \cos(\bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta} \\ &= \sum_{k=1}^n \int \bar{a}_k(\bar{\theta}) \cos(\theta_k(t) + \bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta} + \sum_{k=1}^n \int \bar{a}_k(\bar{\theta}) \cos(\theta_k(t) - \bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta} \\ &= \sum_{k=1}^n \int \bar{a}_k(\bar{\theta}) \cos(\theta_k(t) + \bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta} + \sum_{k \neq k_0} \int \bar{a}_k(\bar{\theta}) \cos(\theta_k(t) - \bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta} \\ &\quad + \int \bar{a}_{k_0}(\bar{\theta}) \cos(\theta_{k_0}(t) - \bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta} \\ &= I + II + III. \end{aligned} \quad (67)$$

For I and II, the scale separation assumption of the data implies that $\bar{a}_k(\bar{\theta}) \cos(\theta_k(t) + \bar{\theta})$ and $\bar{a}_k(\bar{\theta}) \cos(\theta_k(t) - \bar{\theta})$ ($k \neq k_0$) are more oscillatory than the kernel $\phi(\bar{\theta})$. Thus we expect that these two terms are small after convolution with a smooth kernel. In fact, we can prove that $I, II = O(\epsilon)$ by Corollary 6.1.

The estimate for III is more difficult. By our assumption, we have $\left| \frac{(\theta_{k_0}^n)'(t)}{\theta_{k_0}'(t)} - 1 \right| < \frac{\Delta}{2}$.

This implies that $\bar{a}_{k_0}(\bar{\theta}) \cos(\theta_{k_0}(t) - \bar{\theta})$ is smoother than the kernel $\phi(\bar{\theta})$. Using Corollary 6.1, we can prove that $III = \bar{a}_{k_0}(\bar{\theta}) \cos(\theta_{k_0}(t) - \bar{\theta}) + O(\epsilon)$. By combining these results, we get

$$a(t) = A(\bar{\theta}(t)) = a_{k_0}(t) \cos(\theta_{k_0}(t) - \bar{\theta}) + O(\epsilon). \quad (68)$$

Similarly, we can prove the estimate for $b(t)$

$$b(t) = B(\bar{\theta}(t)) = a_{k_0}(t) \sin(\theta_{k_0}(t) - \bar{\theta}) + O(\epsilon). \quad (69)$$

This completes the proof. \square

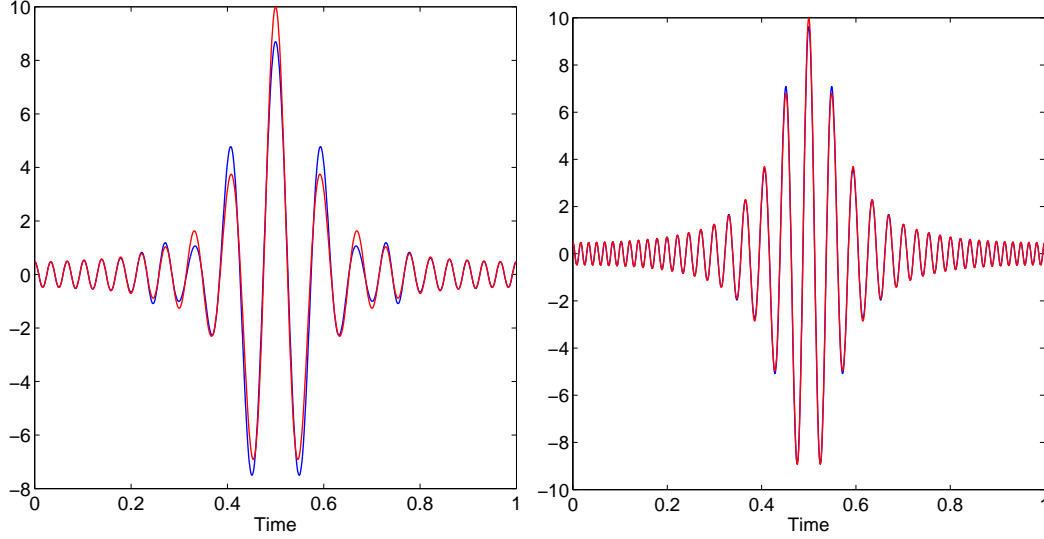


Figure 17: IMFs with different factors of scale separation. Left: poor scale separation; Right: good scale separation.

Remark 6.2 *Under the same assumption, we can prove that the error of the instantaneous frequency is also order ϵ , i.e.*

$$\left| \left(\theta_{k_0}^{n+1} \right)'(t) - \theta'_{k_0}(t) \right| = O(\epsilon). \quad (70)$$

The argument is almost the same as the above proof, except that the calculation is a little more involved.

From the above theorem, we can see that the accuracy of our method depends on the factor of scale separation. This is also consistent with our numerical results. In the following numerical example, we compare the IMFs obtained by our method for two different signals. One has poor scale separation, the other one has better scale separation. The signals are given by (71). The signal f_2 has a better scale separation property than f_1 since its instantaneous frequency is twice of that of f_1 . As shown in Fig. 17, the error we obtain for f_2 is considerably smaller than that of f_1 .

$$\begin{aligned} a_0(t) &= a_1(t) = \frac{1}{1.1 + \cos(2\pi t)}, \quad \theta = 10 \sin(2\pi t) + 40\pi t. \\ f_1(t) &= a_0(t) + a_1(t) \cos \theta(t), \quad f_2(t) = a_0(t) + a_1(t) \cos(2\theta(t)). \end{aligned} \quad (71)$$

We would like to point out that the error estimate given by Theorem 6.1 is highly over-estimated. In Fig. 17, we can see that even for the signal with a poor scale separation property, the instantaneous frequency we obtain is still reasonably accurate, although the corresponding scale separation factor $\epsilon \approx 1.4$ is quite big according to Definition 6.1).

In the estimate of Lemma 6.1, instead of taking the supreme over \mathbb{R} , we can take supreme over a finite interval, since we can choose a low-pass filter ϕ that decays exponentially fast. Then, we can get a more local estimate:

$$\left| \int a(\tau) e^{i\theta(\tau)} \phi(\tau - t) d\tau - a(t) e^{i\theta(t)} \widehat{\phi}(\theta'(t)) \right| \leq \sup_{t \in S_\phi} |a'(t)| I_1 + \frac{1}{2} |a(t)| \sup_{t \in S_\phi} |\theta''(t)| I_2, \quad (72)$$

where $S_\phi = \{t \in \mathbb{R} : |\phi(t)| > \epsilon\}$.

This seems to suggest that for the signal that does not have a good scale separation property in some region, its influence on the accuracy of the decomposition is limited to that region. This is the temporal locality property we have mentioned in Section 5.1. This property can be also seen in Fig. 17. In this example, the scales are not well separated in the center of the interval. However, the scales are better separated near the two ends. We can see that the error near the boundary is much smaller than that in the center.

From this analysis, we can also see that the low-pass filter with smooth Fourier spectrum (such as the cosine function given by (31)) would perform better than that with discontinuous spectrum (such as the stair function given by (30)) in terms of maintaining the temporal locality property of the decomposition. The low-pass filter with discontinuous spectrum decays much slower in the time domain due to the Gibbs phenomena. This is why we use the cosine low-pass filter instead of the stair one.

Theorem 6.1 tells us that if we have a good initial guess, then there is no need to do iterations. But in most cases, we have only a rough initial guess, and the condition in Theorem 6.1 may not be satisfied. In this case, the iterative procedure in our algorithm improves the result gradually as the number of iterations increases. It will provide a good approximation after a number of iterations. In Fig. 18, we show how the iteration can improve the approximation of the instantaneous frequency for a simple chirp signal: $f(t) = \cos(10\pi(3t + 1)^2)$.

In this example, the initial guess for the instantaneous frequency is a constant, $\theta_0 = 80\pi t$. Near the intersection of $\theta'_0 = 80\pi$ and the exact instantaneous frequency, the initial guess is relatively good. After one step, we can get a better approximation in this local region. This new estimate gives us a better guess for the instantaneous frequency in a slightly larger interval that contains the good interval given by the initial guess. Then in the next step, we can get a good approximation in an even larger interval. Gradually, we can get an accurate approximation of the instantaneous frequency in the whole interval. Fig. 18 plots the approximate instantaneous frequency in different steps. As the number of iterations increases, the region in which we have a good approximation becomes larger and larger. Finally, the iterative algorithm produces an accurate instantaneous frequency in the entire domain.

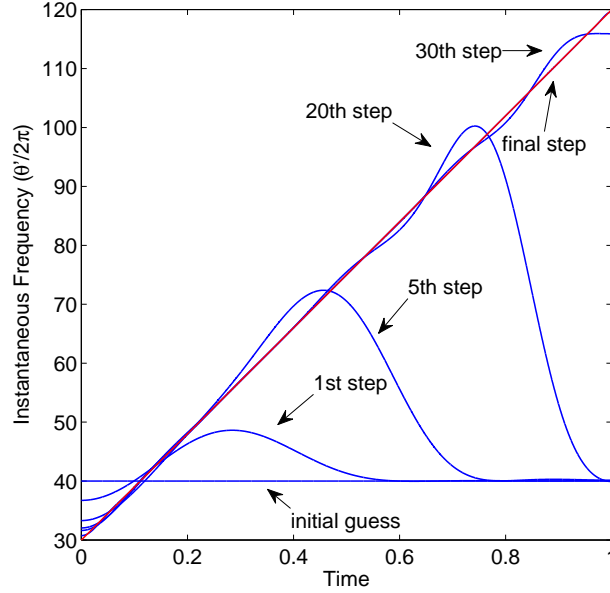


Figure 18: Instantaneous frequency at different step.

7 Conclusion

In this paper, we introduce a new data-driven time-frequency analysis method based on the nonlinear matching pursuit. The adaptivity of our decomposition is obtained by looking for the sparsest representation of signals in the time-frequency domain from a largest possible dictionary that consists of all possible candidates for intrinsic mode functions (IMFs). Solving this nonlinear optimization problem is in general very difficult. We propose a nonlinear matching pursuit method to solve this nonlinear optimization problem by generalizing the classical matching pursuit for the l^0 optimization problem. One important advantage of this nonlinear matching pursuit method is it can be implemented very efficiently. Further, this approach is very stable to noise. For data with good scale separation property, our method gives an accurate decomposition up to the boundary.

We have also carried out some preliminary theoretical study for the nonlinear optimization method proposed in this paper. In the case when the signal satisfies certain scale separation conditions, we show that our iterative algorithm converges to an approximate decomposition with the accuracy determined by the scale separation factor of the signal.

There are some remaining issues to be studied in the future, such as data with poor scale separation property, data with intra-wave frequency modulation, the so-called 'end effect' of data, and data with incomplete or sparse samples and so on. We have addressed these issues to some extent in this paper, but much more work need to be done to resolve these challenging issues.

Another direction is to generalize this adaptive data analysis method to high dimensional data. In some physical applications such as propagation of nonlinear ocean waves, each wave form has a dominating propagation direction. In this case, our method has a natural generalization by adopting a multi-dimensional phase function.

Appendix

Proof of Theorem 6.1: In order to simplify the notation, we denote $\bar{\theta} = \theta_{k_0}^n$ and use $\overline{(\cdot)}$ to represent the mapping from t to $\bar{\theta}$, i.e. $\bar{f}(\bar{\theta}) = f(t)$, $\forall f$.

According to our algorithm, we update $\theta_{k_0}^{n+1}$ as follows:

$$\theta_{k_0}^{n+1} = \bar{\theta} - \arctan\left(\frac{b(t)}{a(t)}\right), \quad a(t) = A(\bar{\theta}(t)), \quad b(t) = B(\bar{\theta}(t)), \quad (73)$$

where

$$A(\gamma) = 2 \int \bar{f}(\bar{\theta}) \cos(\bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta}, \quad B(\gamma) = 2 \int \bar{f}(\bar{\theta}) \sin(\bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta}. \quad (74)$$

We first estimate $A(\gamma)$ as follows:

$$A(\gamma) = 2 \int \bar{f}(\bar{\theta}) \cos(\bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta} = 2 \sum_{k=1}^n \int \bar{a}_k(\bar{\theta}) \cos \theta_k(t) \cos(\bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta}.$$

For $k \neq k_0$, we have

$$\begin{aligned} & 2 \int \bar{a}_k(\bar{\theta}) \cos \theta_k(t) \cos(\bar{\theta}) \phi(\bar{\theta} - \gamma) d\bar{\theta} \\ &= \int \bar{a}_k(\bar{\theta}) (\cos(\theta_k(t) + \bar{\theta}) + \cos(\theta_k(t) - \bar{\theta})) \phi(\bar{\theta} - \gamma) d\bar{\theta}. \end{aligned} \quad (75)$$

Since

$$\left| \frac{d\bar{a}_k(\bar{\theta})}{d\bar{\theta}} \right| = \left| \frac{a'_k(t)}{\bar{\theta}'(t)} \right| \leq \epsilon \left| \frac{\theta'_k(t)}{\bar{\theta}'(t)} \right|, \quad (76)$$

we obtain

$$\left| \frac{d^2}{d\bar{\theta}^2} (\theta_k(t) \pm \bar{\theta}) \right| = \left| \frac{\theta''_k(t) \bar{\theta}'(t) - \theta'_k(t) \bar{\theta}''(t)}{(\bar{\theta}'(t))^3} \right| \leq \epsilon \left| \frac{(\theta'_k(t))^2 - \theta'_k(t) \bar{\theta}'(t)}{(\bar{\theta}'(t))^2} \right|, \quad \text{for } k \neq k_0. \quad (77)$$

Now we apply Corollary 6.1 for $k \neq k_0$ to obtain

$$\begin{aligned} & 2 \int \bar{a}_k(\bar{\theta}) \cos \theta_k(t) \cos(\bar{\theta}) \phi(\bar{\theta} - \phi) d\bar{\theta} \\ &= \bar{a}_k(\bar{\theta}) \cos(\theta_k(t) + \bar{\theta}) \hat{\phi}\left(\frac{\theta'_k(t)}{\bar{\theta}'(t)} + 1\right) + \bar{a}_k(\bar{\theta}) \cos(\theta_k(t) - \bar{\theta}) \hat{\phi}\left(\frac{\theta'_k(t)}{\bar{\theta}'(t)} - 1\right) + O(\epsilon). \end{aligned} \quad (78)$$

Using the condition $\Delta < \frac{d-1}{d+1/2}$, $\theta'_k(t) > d\theta'_{k-1}(t)$ and $\left| \frac{\bar{\theta}'(t)}{\theta'_{k_0}(t)} - 1 \right| < \frac{\Delta}{2}$, we get

$$\frac{\theta'_k(t)}{\bar{\theta}'(t)} - 1 = \frac{\theta'_k(t)}{\theta_{k_0}(t)} \frac{\theta'_{k_0}(t)}{\bar{\theta}'(t)} - 1 > d(1 - \Delta/2) - 1 > \Delta, \quad \text{if } k > k_0, \quad (79)$$

$$\frac{\theta'_k(t)}{\bar{\theta}'(t)} - 1 = \frac{\theta'_k(t)}{\theta_{k_0}(t)} \frac{\theta'_{k_0}(t)}{\bar{\theta}'(t)} - 1 < (1 + \Delta/2)/d - 1 < -\Delta, \quad \text{if } k < k_0, \quad (80)$$

$$\frac{\theta'_k(t)}{\bar{\theta}'(t)} + 1 > 1 > \Delta. \quad (81)$$

Since the support of $\hat{\phi}$ is within $[-\Delta, \Delta]$, we have

$$2 \int \bar{a}_k(\bar{\theta}) \cos \theta_k(t) \cos(\bar{\theta}) \phi(\bar{\theta} - \phi) d\bar{\theta} = O(\epsilon). \quad (82)$$

For $k = k_0$, we proceed as follows

$$\begin{aligned} & 2 \int \bar{a}_k(\bar{\theta}) \cos \theta_k(t) \cos(\bar{\theta}) \phi(\bar{\theta} - \phi) d\bar{\theta} \\ &= \int \bar{a}_{k_0}(\bar{\theta}) (\cos(\theta_{k_0}(t) + \bar{\theta}) + \cos(\theta_{k_0}(t) - \bar{\theta})) \phi(\bar{\theta} - \phi) d\bar{\theta}. \end{aligned} \quad (83)$$

Similarly, by using the assumption

$$\left| \frac{d\bar{a}_{k_0}(\bar{\theta})}{d\bar{\theta}} \right| = \left| \frac{a'_{k_0}(t)}{\bar{\theta}'(t)} \right| \leq \epsilon \left| \frac{\theta'_{k_0}(t)}{\bar{\theta}'(t)} \right|, \quad (84)$$

we obtain the following estimates:

$$\left| \frac{d}{d\bar{\theta}} (\theta_{k_0}(t) + \bar{\theta}) \right| = \left| \frac{\theta'_{k_0}(t)}{\bar{\theta}'(t)} + 1 \right| > 1 > \Delta, \quad (85)$$

$$\left| \frac{d}{d\bar{\theta}} (\theta_{k_0}(t) - \bar{\theta}) \right| = \left| \frac{\theta'_{k_0}(t)}{\bar{\theta}'(t)} - 1 \right| < \frac{\Delta}{2}, \quad (86)$$

$$\left| \frac{d^2}{d\bar{\theta}^2} (\theta_{k_0}(t) \pm \bar{\theta}) \right| = \left| \frac{\theta''_{k_0}(t) \bar{\theta}'(t) - \theta'_{k_0}(t) \bar{\theta}''(t)}{(\bar{\theta}'(t))^3} \right| \leq \epsilon \left| \frac{(\theta'_{k_0}(t))^2 - \theta'_{k_0}(t) \bar{\theta}'(t)}{(\bar{\theta}'(t))^2} \right|. \quad (87)$$

By applying Corollary 6.1 again, we get

$$\begin{aligned} & 2 \int \bar{a}_k(\bar{\theta}) \cos \theta_k(t) \cos(\bar{\theta}) \phi(\bar{\theta} - \phi) d\bar{\theta} \\ &= \bar{a}_{k_0}(\bar{\theta}) \cos(\theta_{k_0}(t) + \bar{\theta}) \hat{\phi} \left(\frac{\theta'_{k_0}(t)}{\bar{\theta}'(t)} + 1 \right) + \bar{a}_{k_0}(\bar{\theta}) \cos(\theta_{k_0}(t) - \bar{\theta}) \hat{\phi} \left(\frac{\theta'_{k_0}(t)}{\bar{\theta}'(t)} - 1 \right) + O(\epsilon) \\ &= \bar{a}_{k_0}(\bar{\theta}) \cos(\theta_{k_0}(t) - \bar{\theta}) + O(\epsilon). \end{aligned} \quad (88)$$

Finally, we get the following estimate for $a(t)$,

$$a(t) = A(\bar{\theta}(t)) = a_{k_0}(t) \cos(\theta_{k_0}(t) - \bar{\theta}) + O(\epsilon). \quad (89)$$

For $b(t)$, we can obtain a similar estimate

$$b(t) = B(\bar{\theta}(t)) = a_{k_0}(t) \sin(\theta_{k_0}(t) - \bar{\theta}) + O(\epsilon). \quad (90)$$

Thus, we have

$$\Delta\theta = \arctan\left(\frac{B(t)}{A(t)}\right) = \theta_{k_0}(t) - \bar{\theta} + O(\epsilon), \quad (91)$$

which implies that

$$\left|\theta_{k_0}^{n+1}(t) - \theta_{k_0}(t)\right| = O(\epsilon). \quad (92)$$

This completes the proof. \square

Acknowledgments.

This work was in part supported by the AFOSR MURI grant FA9550-09-1-0613. We would like to thank Professors Norden E. Huang and Zhaohua Wu for many stimulating discussions on EMD/EEMD and topics related to the research presented here. We would also like to thank Professors Ingrid Daubechies, Stanley Osher, and Zuowei Shen for their interest in this work and for a number of valuable discussions. Prof. Hou would like to express his gratitude to the National Central University (NCU) for their support and hospitality during his visits to NCU in the past two years.

References

- [1] B. Boashash, Time-Frequency Signal Analysis: Methods and Applications, Longman-Cheshire, Melbourne and John Wiley Halsted Press, New York, 1992.
- [2] A. M. Bruckstein, D. L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM Review*, **51**, pp. 34-81, 2009.
- [3] E. Candès and T. Tao, Near optimal signal recovery from random projections: Universal encoding strategies?, *IEEE Trans. on Information Theory*, **52(12)**, pp. 5406-5425, 2006.
- [4] E. Candes, J. Romberg, and T. Tao, Robust uncertainty principles: Exact signal recovery from highly incomplete frequency information, *IEEE Trans. Inform. Theory*, **52**, pp. 489-509, 2006.
- [5] S. Chen, D. Donoho and M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, **20**, pp. 33-61, 1998.

- [6] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series on Applied Mathematics, Vol. 61, SIAM Publications, 1992.
- [7] I. Daubechies, J. Lu and H. Wu, Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool, *Appl. Comp. Harmonic Anal.*, **30** (2011), pp. 243-261.
- [8] D. L. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory*, **52**, pp. 1289-1306, 2006.
- [9] P. Flandrin, Time-Frequency/Time-Scale Analysis, Academic Press, San Diego, CA, 1999.
- [10] R. S. Gross, 2001 Combinations of Earth orientation measurements: SPACE2000, COMB2000, and POLE2000. JPL Publication 01-2. Jet Propulsion Laboratory, Pasadena, CA.
- [11] D. Gabor, Theory of communication, *J. IEE.*, **93**, pp. 426-457, 1946.
- [12] Tom Goldstein and Stanley Osher, The Split Bregman Method for L_1 -Regularized Problems, *SIAM J. Imaging Sci.*, **2**, pp. 323-343, 2009.
- [13] T. Y. Hou and Z. Shi, Adaptive Data Analysis via Sparse Time-Frequency Representation, *Advances in Adaptive Data Analysis*, **3**, pp. 1-28, 2011
- [14] N. E. Huang et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. Lond. A*, **454** (1998), pp. 903-995.
- [15] D. L. Jones and T. W. Parks, A high resolution data-adaptive time-frequency representation, *IEEE Trans. Acoust. Speech Signal Process*, **38**, pp. 2127-2135, 1990.
- [16] P. J. Loughlin and B. Tracer, On the amplitude - and frequency-modulation decomposition of signals, *J. Acoust. Soc. Am.*, **100**, pp. 1594-1601, 1996.
- [17] B. C. Lovell, R. C. Williamson and B. Boashash, The relationship between instantaneous frequency and timefrequency representations, *IEEE Trans. Signal Process*, **41**, pp. 1458-1461, 1993.
- [18] S. Mallat and Z. Zhang, Matching pursuit with time-frequency dictionaries, *IEEE Trans. Signal Process*, **41**, pp. 3397-3415, 1993.
- [19] S. Mallat, A wavelet tour of signal processing: the Sparse way, Academic Press, 2009.
- [20] W. K. Meville, Wave modulation and breakdown, *J. Fluid Mech.*, **128**, pp. 489-506, 1983.
- [21] S. Olhede and A. T. Walden, The Hilbert spectrum via wavelet projections, *Proc. Roy. Soc. London A*, **460**, pp. 955-975, 2004.

- [22] B. Picinbono, On instantaneous amplitude and phase signals, *IEEE Trans. Signal Process.*, **45** (1997), pp. 552-560.
- [23] S. Qian and D. Chen, Joint Time-Frequency Analysis: Methods and Applications, Prentice Hall, 1996.
- [24] S. Qian and D. Chen, Signal representation using adaptive normalized Gaussian functions, *Signal Processing*, **36** pp. 1-11, 1994.
- [25] S. O. Rice, Mathematical analysis of random noise, *Bell Syst. Tech. J.*, **23**, pp. 282-310, 1944.
- [26] J. Shekel, Instantaneous frequency, *Proc. IRE*, **41** , pp. 548-548, 1953.
- [27] J. Tropp and A. Gilbert, Signal recovery from random measurements via Orthogonal Matching Pursuit, *IEEE Trans. Inform. Theory*, **53**, pp. 4655-4666, 2007.
- [28] B. Van der Pol, The fundamental principles of frequency modulation, *Proc. IEE*, **93**, pp. 153-158, 1946.
- [29] Hau-tieng Wu, Instantaneous frequency and wave shape functions (I), arXiv:1104.2365v1.
- [30] Z. Wu and N. E. Huang, Ensemble Empirical Mode Decomposition: a noise-assisted data analysis method, *Advances in Adaptive Data Analysis*, **1**, pp. 1-41, 2009.